Math.--Econ.
Library
HA
29
.K12

B 456085

STATISTICS

WITHOUT NUMBERS

A Visual Explanation

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

PROPERTY OF
University of Michigan Libraries
1817
ARTES SCIENTIA VERITAS

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

S 2^{00}

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

STATISTICS WITHOUT NUMBERS

A Visual Explanation

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

STATISTICS WITHOUT NUMBERS

A Visual Explanation

By

F. KAFKA, J.D., Pol.Sc.D.
The City College, New York

LIFETIME EDITIONS

NEW YORK

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Math.-Econ.
Library
HA
29
.K12

Copyright, 1950
By F. Kafka

PRINTED IN THE UNITED STATES OF AMERICA
BY J. J. LITTLE & IVES COMPANY, NEW YORK

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

13- ...

Contents

Preface	ix
What Statistics Can Do and What It Cannot Do	13
Ratios	17
Frequency Distribution	25
Time Series	49
Index Numbers	63
Correlation	77
Sampling	89
The Normal Distribution	95
References	103
Index	109

Digitized by Google
Original from UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

If this volume attains its goal of clarifying statistics to the layman, it is largely due to the artistic ability and ingenuity of Mr. E. Phillips. It was he who helped to make it a book of statistics without numbers.

The text benefited from the helpful suggestions offered by Mr. S. Dutka, statistician with Elmo Roper, New York, and Mr. B. Lipstein, statistician with the Bureau of Labor Statistics, Washington.

Digitized by Google Original from UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Preface

THIS BOOK is addressed to the consumer of statistics. It is meant for the newspaper reader who every day may come across statistical terms in his paper, for the businessman who finds statistical terms in many reports, for the secretary who has to use them frequently and generally, for the alert person who has not received training in statistics.

It is not designed for the reader trained in statistics and does not intend to compete with any textbook on that subject. It tries to give a few basic statistical concepts to those who otherwise would be left without any tools for interpreting statistics.

The limitations of statistics are emphasized. No attempt is made to explain the computation of statistical measures. Graphic presentation is not covered since it does not require simplification.

The main fields to which this volume applies are economics and business. However, since statistics is presented here with the aid of examples, it is hoped that this work will be useful in a more general way.

In explaining statistical concepts, visual presenta-

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

tion has taken the place of numerical illustration. Basic statistical concepts are presented in rough sketch form, not as precise photographs.

Oversimplification is the price that had to be paid for a discussion of this type.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

STATISTICS WITHOUT NUMBERS

A Visual Explanation

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

What Statistics Can Do and What It Cannot Do

IF ALL PRICES, all wages, all sales figures, all height data, all IQ's, etc., were alike there would be little need for statistics. There is, however, variation everywhere. This necessitates comparisons. The main task of statistics is to help make comparisons.

To make comparisons possible or at least to facilitate them, the raw material of statistical work, the data, have to be organized.

We may just organize the data by classifying or grouping them. The data themselves remain unchanged. A second way is to simplify the data. Here they are changed but their number remains the same. Finally, we may condense our data, thereby changing their quality and number.

After the data have been organized, comparisons can easily be made. We may make comparisons between times or places, with higher categories or

Digitized by Google | Original from UNIVERSITY OF MICHIGAN

with established standards, we may even, in an attempt to forecast, compare the unknown future with the present and the past.

Many objections to statistical work are directed against bad statistics. They criticize the wrong application of statistics, not statistics itself. However, there are significant limitations attached even to good statistics.

Statistical method is based upon the concepts of averages, approximations, and probabilities. Mathematical precision is not to be expected in statistical statements. Furthermore, statistical findings cannot be more perfect or more precise than the data upon which they are based.

Not all statistical results which are stated with great mathematical precision are really that accurate. The use of precise formulae occasionally obscures the shortcomings of data and methods. Precise looking statistical results may be misleading; they may have been arrived at with fairly rough methods.

Statistics depends on the statistician to a great extent. The subjective element is very strong and the work of the statistician very often reflects his personality. Statistics certainly has some of the elements of an art.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The reaction of the public to statistics frequently runs to extremes. On the one hand, we find a dogmatic approach: every printed word, and, especially, every printed figure, is backed by a presumption of truth. On the other hand, we find a large group of people who are skeptical of any statistical presentation.

It seems that, as happens so often, the truth lies somewhere between two extremes. A critical attitude is the best one. It accepts statistical findings gratefully where they supply information which is otherwise not available or not as concisely presented. However, it views statistical results in the light of all limitations attached to statistical work.

With these guide lines in mind, we may now turn to our first statistical concept.

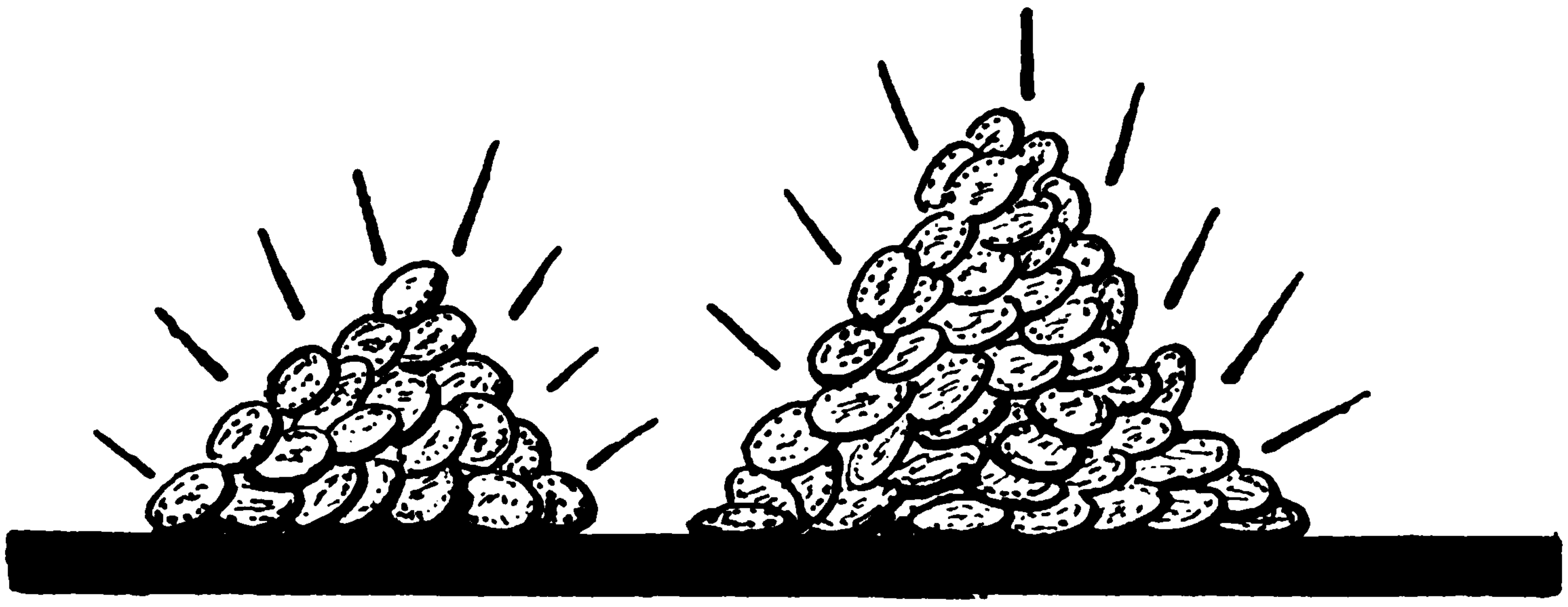
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Ratios

Two retail stores of the Alpha Corporation turn in the cash received during one day.



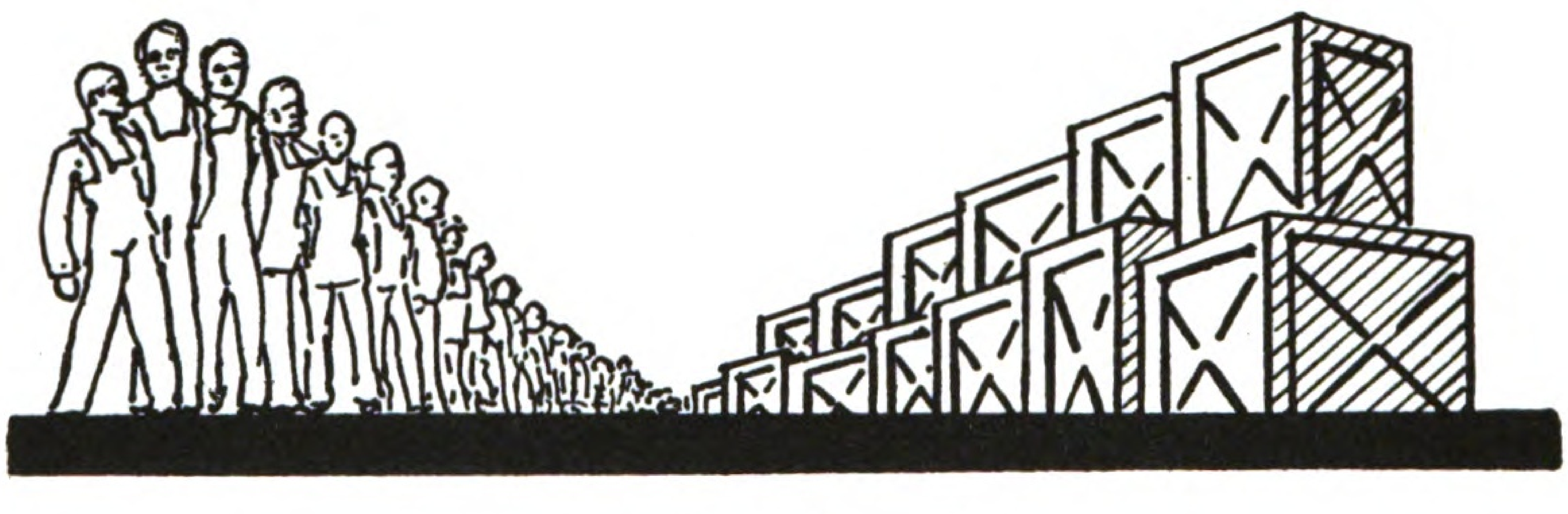
We are not interested in the actual amount of money for the moment but we would like to compare the sales performance of the two stores. We would like to make the simplest possible comparison between these two piles of money.

We find that every Dollar in the small pile can be matched with four Dollars in the large one.

Digitized by Google Original from UNIVERSITY OF MICHIGAN

That means, the cash receipts of the two stores compare as one and four.

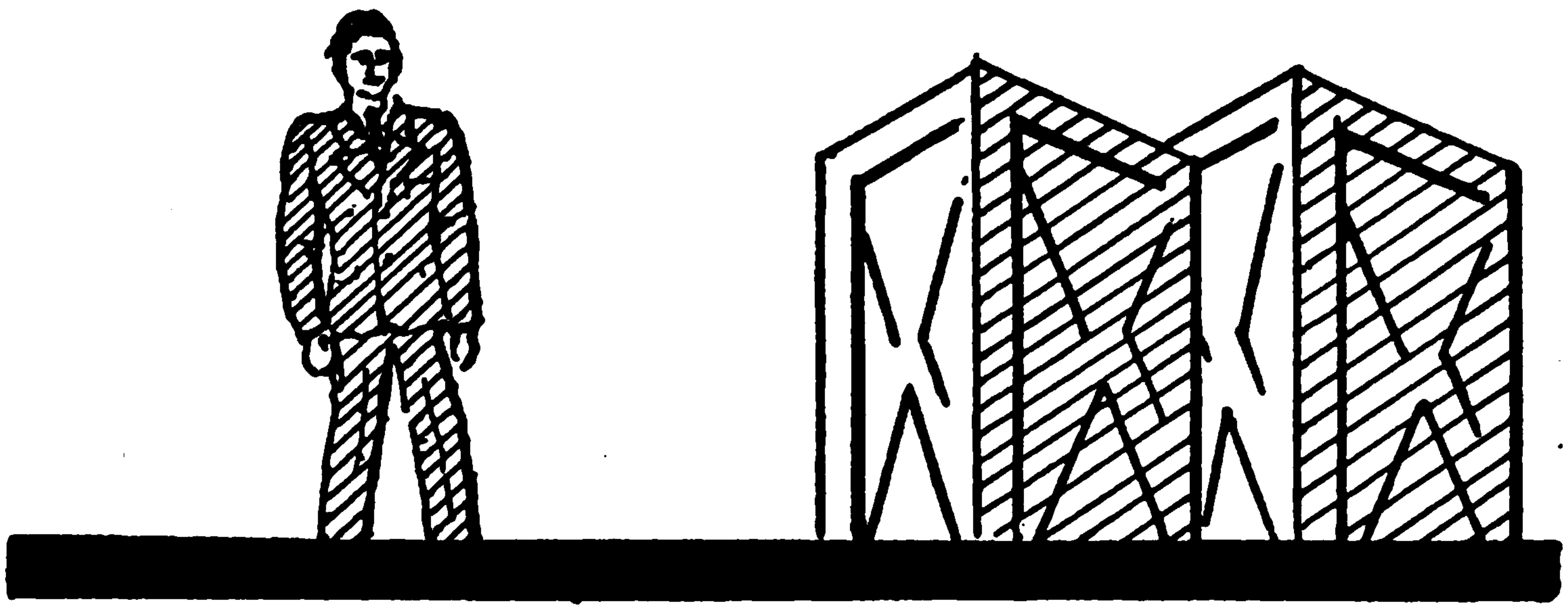
This simplified comparison of the two piles of money is a ratio. We do not have to compare quantities of the same kind. We also call it a ratio if we compare, for instance, the total consumption of a commodity in a city and the total number of inhabitants.



Digitized by Google

Original from UNIVERSITY OF MICHIGAN

We may match every inhabitant with a certain number of boxes. That number shows us the consumption "by heads," or the per capita ratio.



Any two quantities may be compared and the relationship expressed as a ratio. However, a ratio is valid only if quantities are compared which are comparable from the standpoint of logic. We would, for instance, not compare the total consumption of a drug for diabetes patients and the total population. There would be no logical basis for such a comparison.

Many comparisons, expressed as ratios, have become standardized. Normal ratios or safe ratios have been established as guide lines.

The accountant of the Alpha Corporation compares total current assets and total current liabilities. This is the "Current Ratio." He may also compare

Digitized by Google

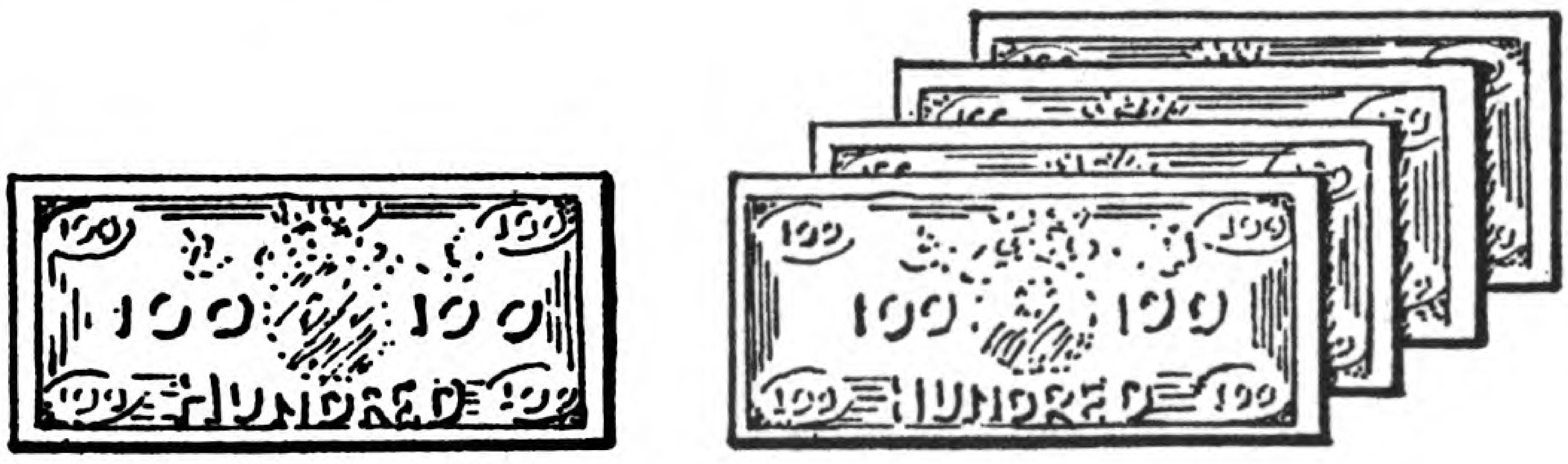
Original from
UNIVERSITY OF MICHIGAN

through the "Net Worth Ratio" net worth and total debts, or through the "Fixed Assets Ratio" net worth and fixed assets. These ratios for the Alpha Corporation will be judged in the light of the standard ratio for the entire industry.

The farmer uses the Corn-Hog Ratio as a guide. This means that he compares the value of 100 lbs. of live hogs and of one bushel of corn. When the ratio is higher than the guide line, the farmer will raise hogs; otherwise he will sell corn.

Intelligence Quotients, Birth Rates and Death Rates are illustrations of widely used ratios.

We matched, before, one Dollar of one pile with a number of Dollars in the other pile. Instead of matching one Dollar we may choose any other number of Dollars. Most frequently, we will examine how many items of one group go with one hundred items of the other. This ratio to one hundred is called a percentage.



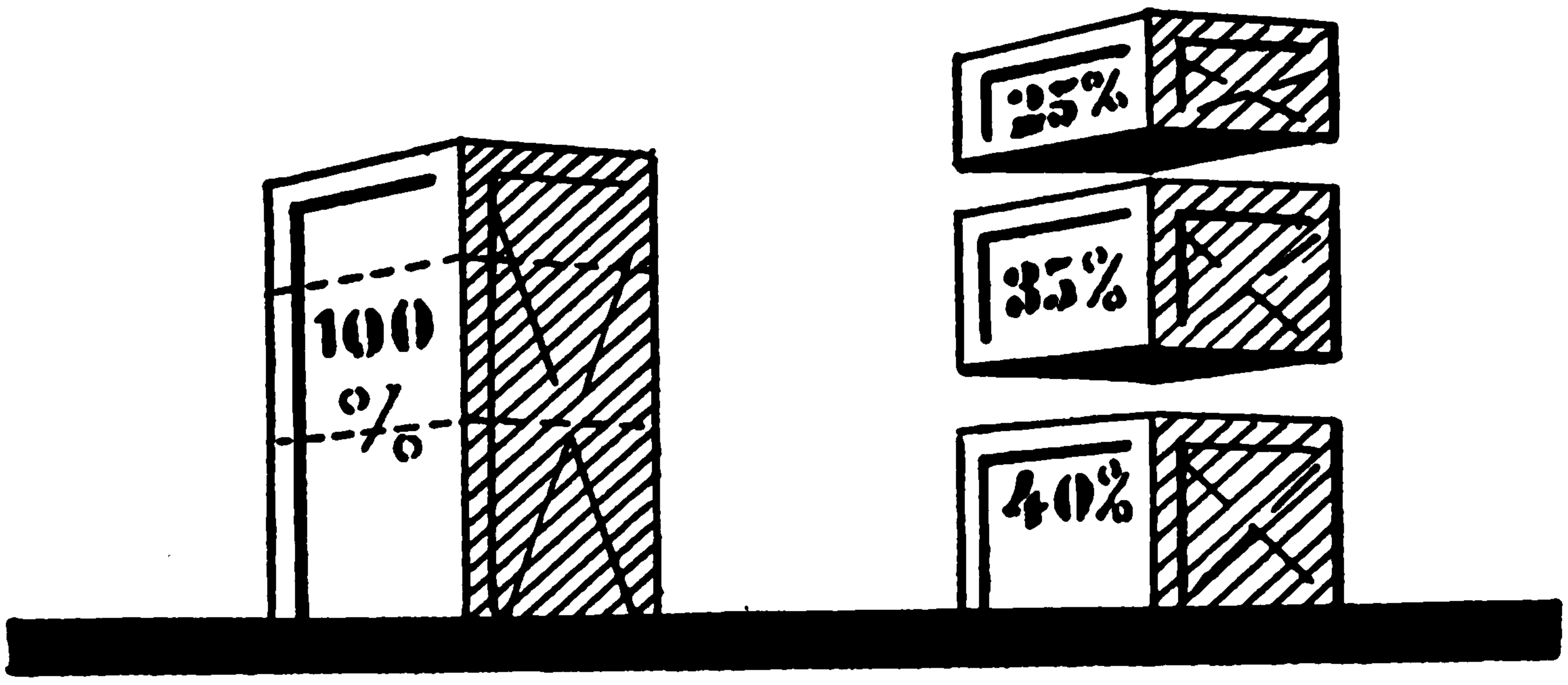
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

This type of ratio should not be used if there are only very few items in both groups. If there are six executives in the Alpha Corporation and two of them are over 60 years, percentages should not be used to express this relationship, since percentages are ratios to 100. This type of ratio furthermore will make comparisons complicated if the items to be compared vary widely. Very high percentage figures make comparisons difficult. The transformation to a ratio to one may facilitate understanding.

We talk of ratios not only when we compare two or more quantities but also when we compare the component parts of a quantity and the total. Ratios, especially percentages, are frequently used in these situations.

For example, the Gamma Corporation manufactures three different articles. We may compare its

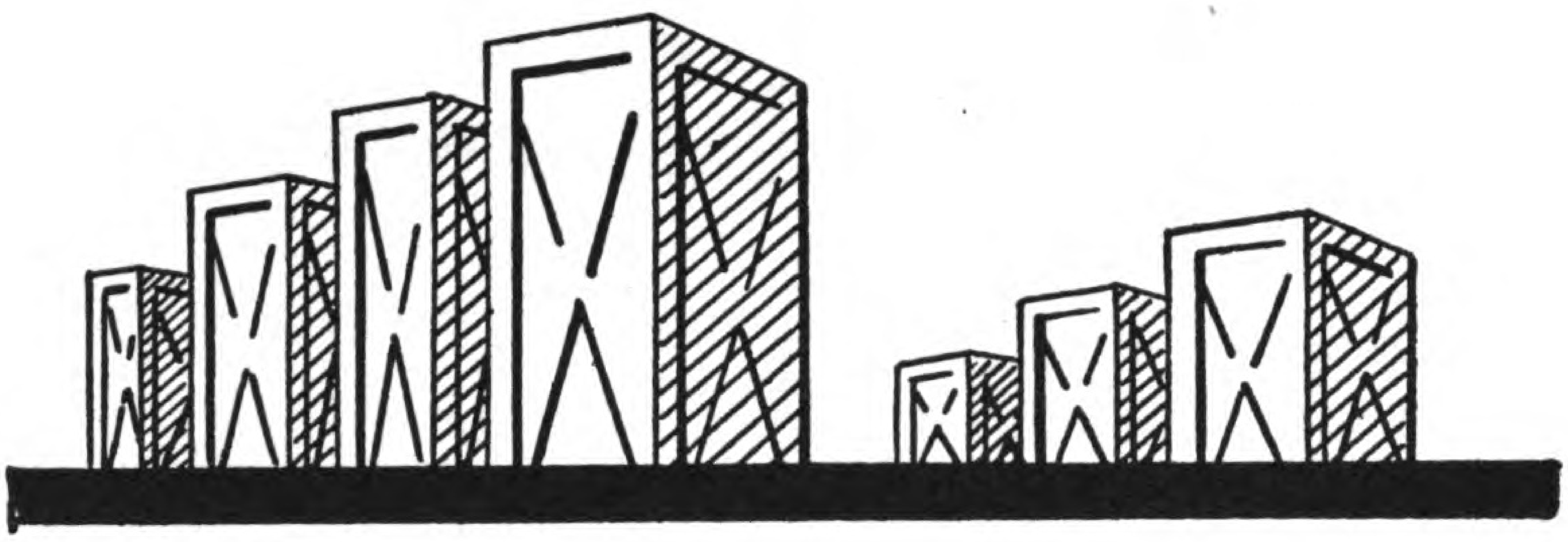


Digitized by Google
Original from UNIVERSITY OF MICHIGAN

total production during one week with the production of the three articles.

The comparison is usually made as a ratio to 100 whereby the total production represents 100 per cent.

Questions may arise about comparison of the 1947 and the 1948 production of the Alpha Corporation.

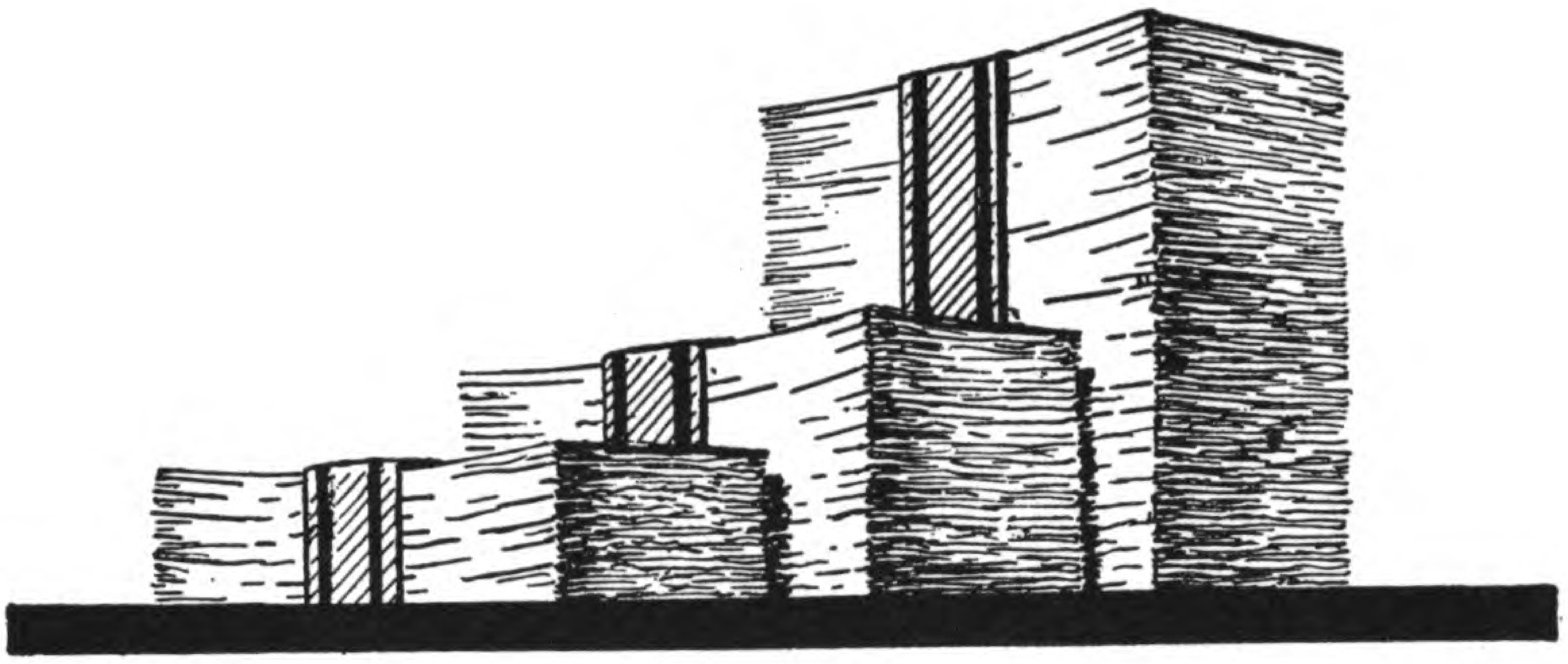


It is equally correct to inquire how much of the 1948 production goes with 100 units of the 1947 production or how much of the 1947 production goes with 100 units of the 1948 production. In other words, either 1947 or 1948 may be made the base of our comparison. We may find the 1948 production expressed in percent of the 1947 production or the 1947 production in percent of the 1948 production. Misunderstandings will invariably result if the base of comparison is not stated clearly.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Here is another cause of errors: A certain amount of money has been received by the Alpha Corporation. This amount can increase 100%, 500%, etc.



What if there is a decrease? A decrease of 50% means a reduction to ½.

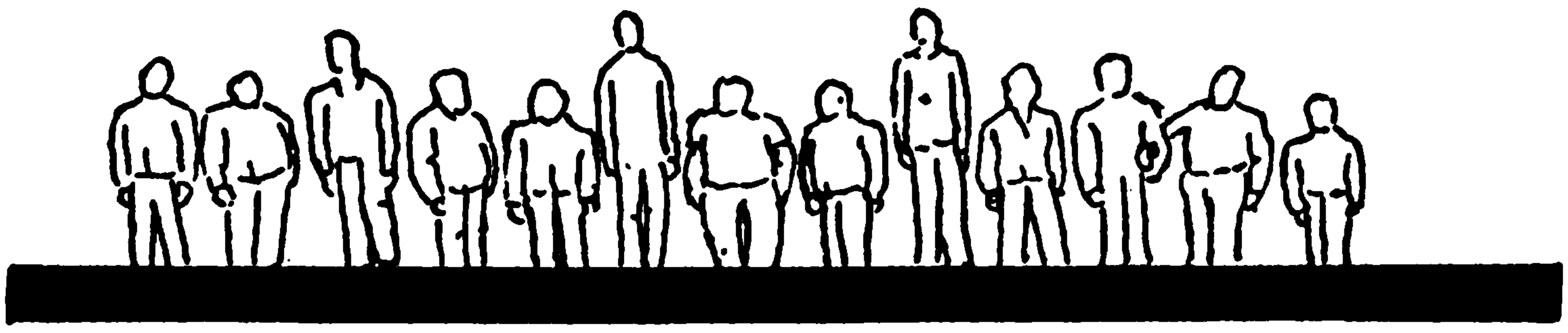
A decrease of 100% means the amount of money has disappeared. A decrease of more than 100% is impossible.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Frequency Distribution

In connection with a health program, the Alpha Corporation intends to study the height of all their male workers.



Obviously, no statement can be made by just looking at these workers or their height data. This is where statistics comes in. It can condense.

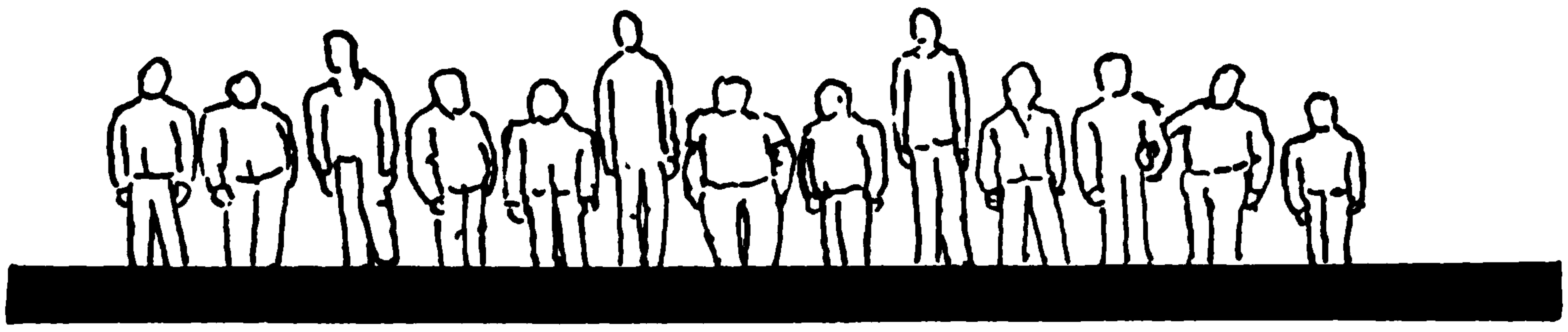
The newspaper editor describes, through use of a headline, the information contained in an article. Let us see whether we can find headlines or shorthand expressions that would describe the height of all Alpha workers in a short and simple way.

There are four different aspects that are capable of measurement and description. This means that if

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

we just mention four characteristic facts, we are adequately describing the height situation of the Alpha workers. By doing so, we have succeeded in condensing this great number of individual height data and we have cleared the path for comparisons.

If we wished to compare for instance the height situation of the male workers with that of the female workers, these four shorthand expressions would enable us to do so. Had we assembled the male and the female workers separately, like this,



no comparison could have been made. We see now that these four shorthand expressions will be of great help.

It is important to note that several different

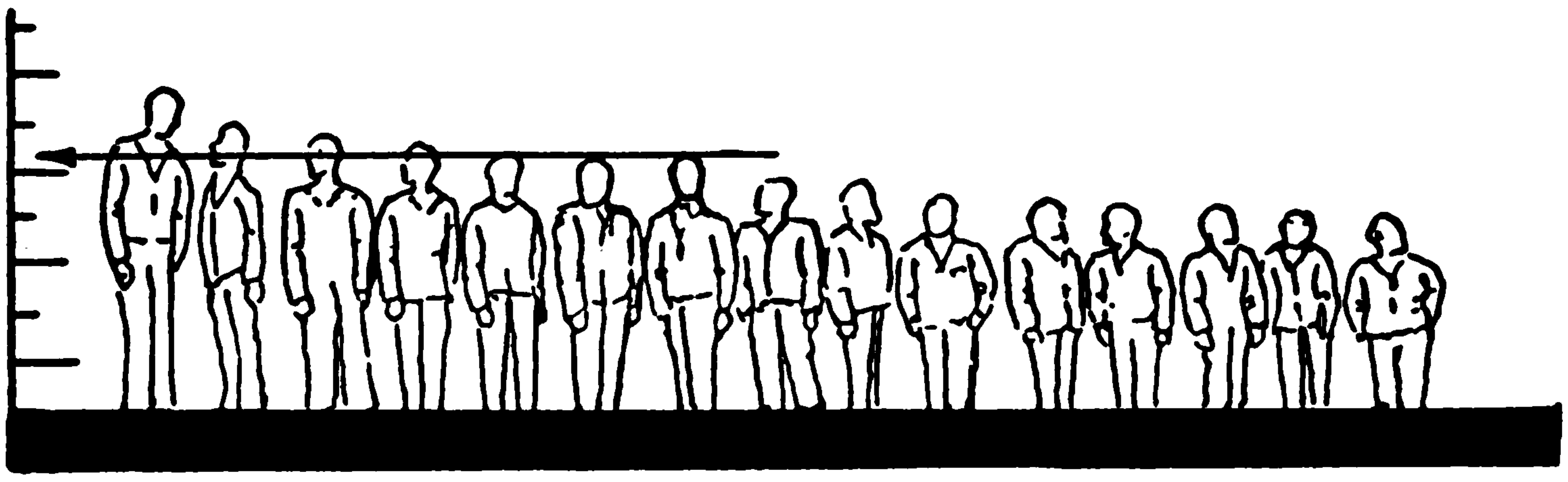
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

"shorthand systems" may be used to write any one of these four shorthand expressions. Thus, there are a number of different measures which may be used to represent the first of these shorthand descriptions, several possibilities for the second shorthand description, and so on.

The First Shorthand Expression

We shall try to find the typical height of the Alpha workers. There are, however, several interpretations of typical height.

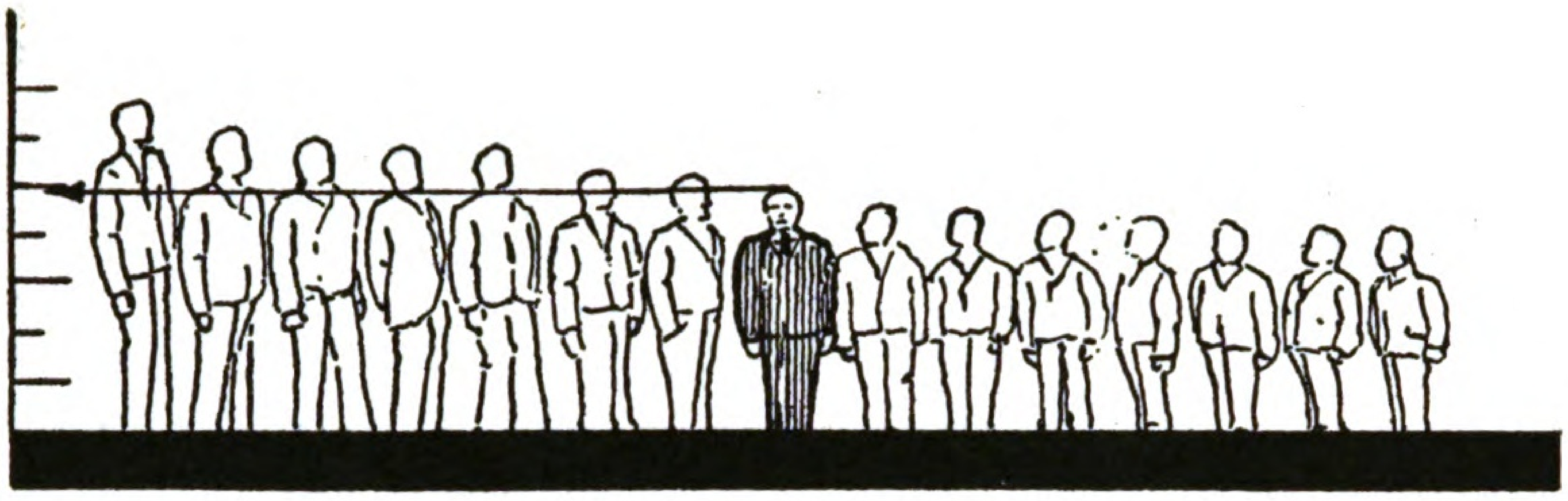
We may consider typical the height which we find most frequently among the workers.



The greatest number of workers has this height, which makes it the most common height. It is the most fashionable height in this group, as it were, and we call it the "mode."

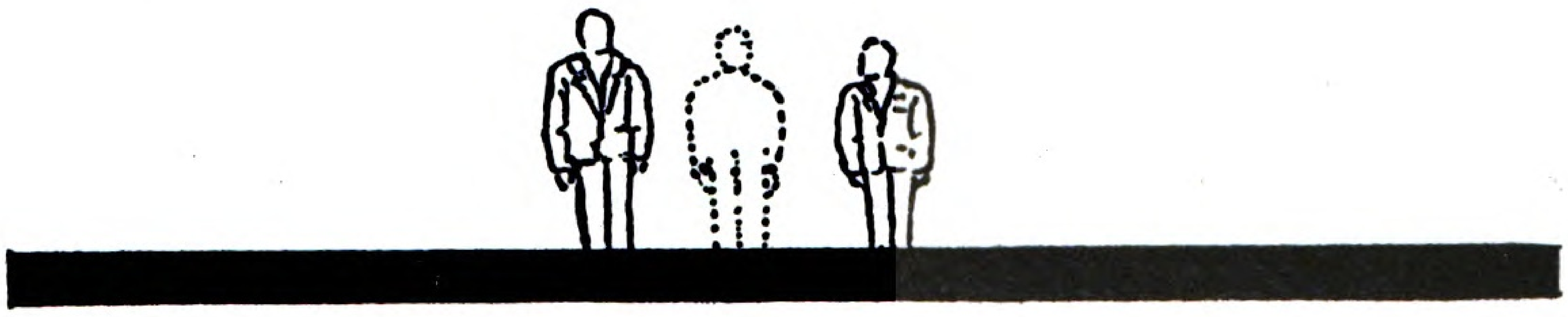
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Another height may be called typical. It is the midmost height.



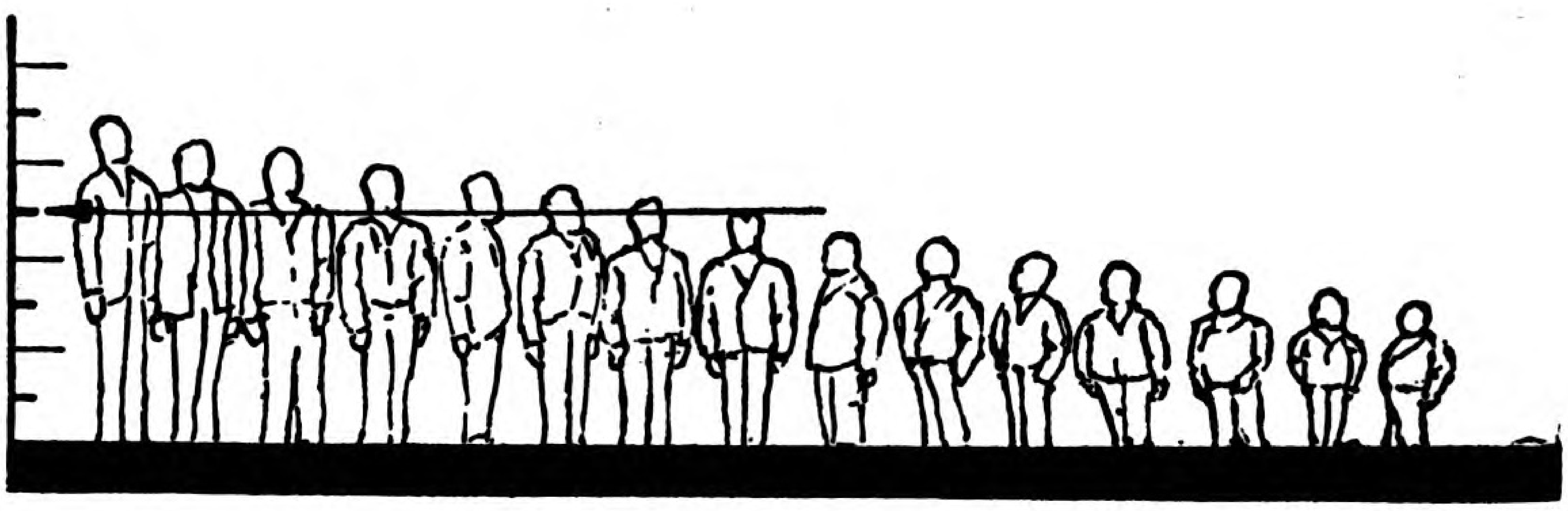
There is an equal number of workers who are taller and shorter than the "middleman." His height is the median. Thus the median divides the number of workers into two equal halves: one half of the workers who are equally tall or taller than the middleman, and one half who are equally tall or shorter than the middleman.

If we had an even number of workers, we would have two middlemen. We would in this case imagine a worker whose height is just halfway between the heights of the two middlemen. The height of this imaginary worker is the median.

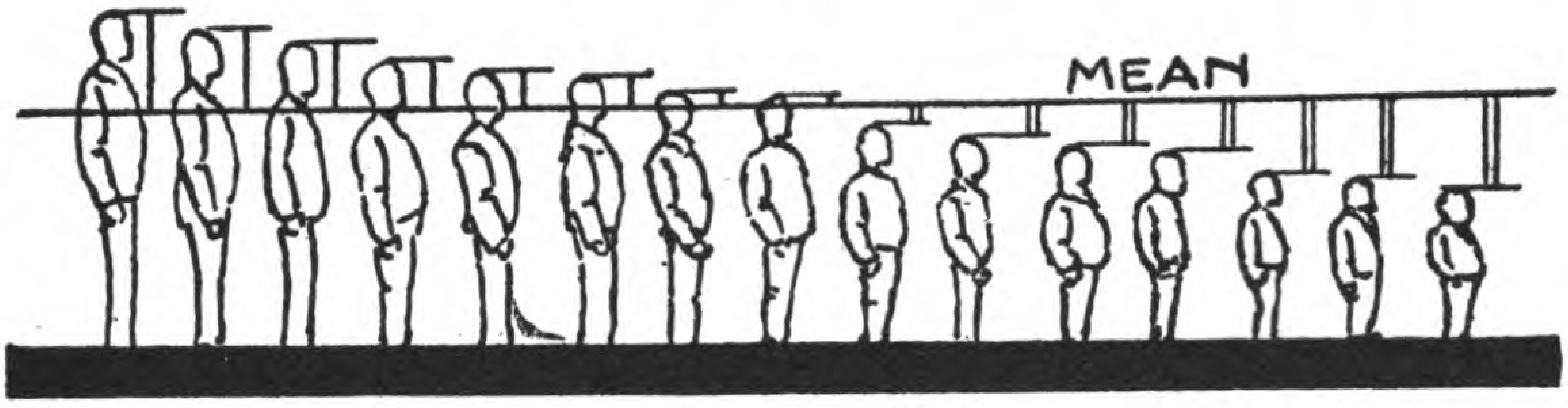


Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The typical height, however, which is best known and most widely used is the arithmetic mean. Many statisticians call it just "the mean" and non-statisticians call it "the average." The mean height is typical because it represents a certain point of balance.



Let us first determine the differences in height between the mean and every single worker.



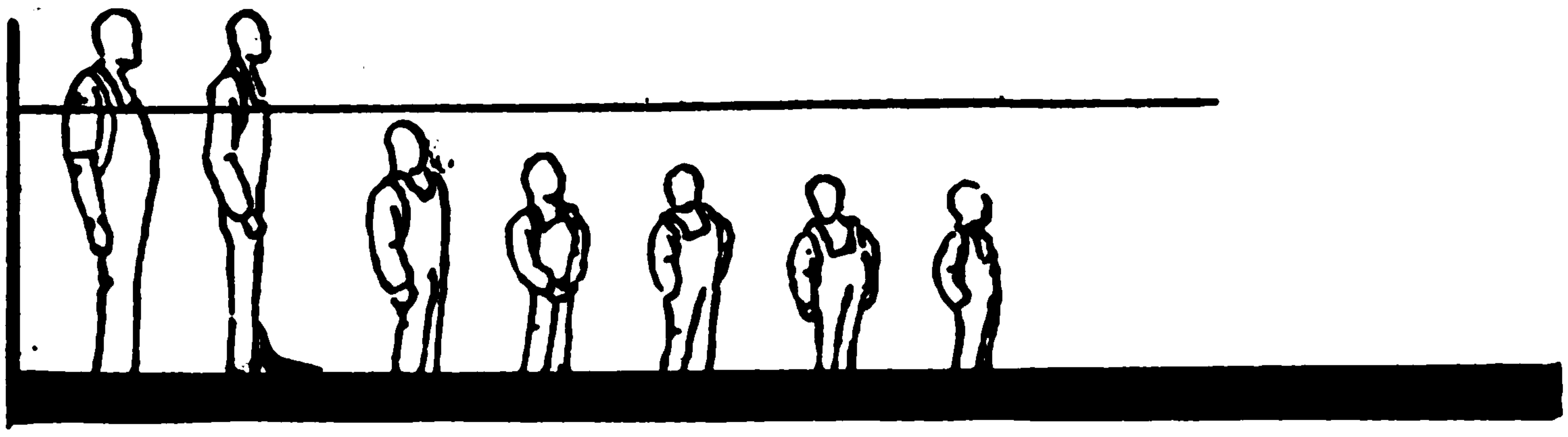
We may now line up all the differences of the workers who are taller than the mean height, and

Digitized by Google
Original from UNIVERSITY OF MICHIGAN

of the workers who are shorter than the mean height.

We see that these two lines are equal. This shows that the mean divides all the workers in such a way that it differs from all taller-than-mean workers exactly as much as from all shorter-than-mean workers.

There are more methods of establishing a typical value, but the arithmetic mean, median, and mode are the most important. Statisticians term averages all these typical values which constitute our first shorthand expression. All averages have their particular usefulness and their particular limitations. For instance, there may not be a significant repetition of the same height; hence there would be no mode. Furthermore, we may have a few extremely tall workers.



Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

This would lift the mean, since the differences to the mean must balance. In such a case, the mean height may cease to be a typical height, as it is influenced by just a few atypically tall—or short—workers. The median or the mode would give a more typical picture, because the height of the middleman or the most frequent height are not affected by the height of the extremely tall or short men.

Averages are frequently used. Let us consider another illustration. The Alpha Corporation is engaged in wage negotiations with a union. Which wage shall be mentioned as typical?

The modal wage is the most common wage. There are more Alpha workers getting this wage than are getting any other rate. The median wage exceeds the modal wage at the Alpha Corporation. Half of all workers receive this median wage or more.

The mean wage, finally, is often referred to as the "average wage." It represents a point of balance and is affected by the magnitude of every single wage as well as its difference to the mean wage. Since there are a few highly paid workers at Alpha, the mean wage is the highest of the three averages.

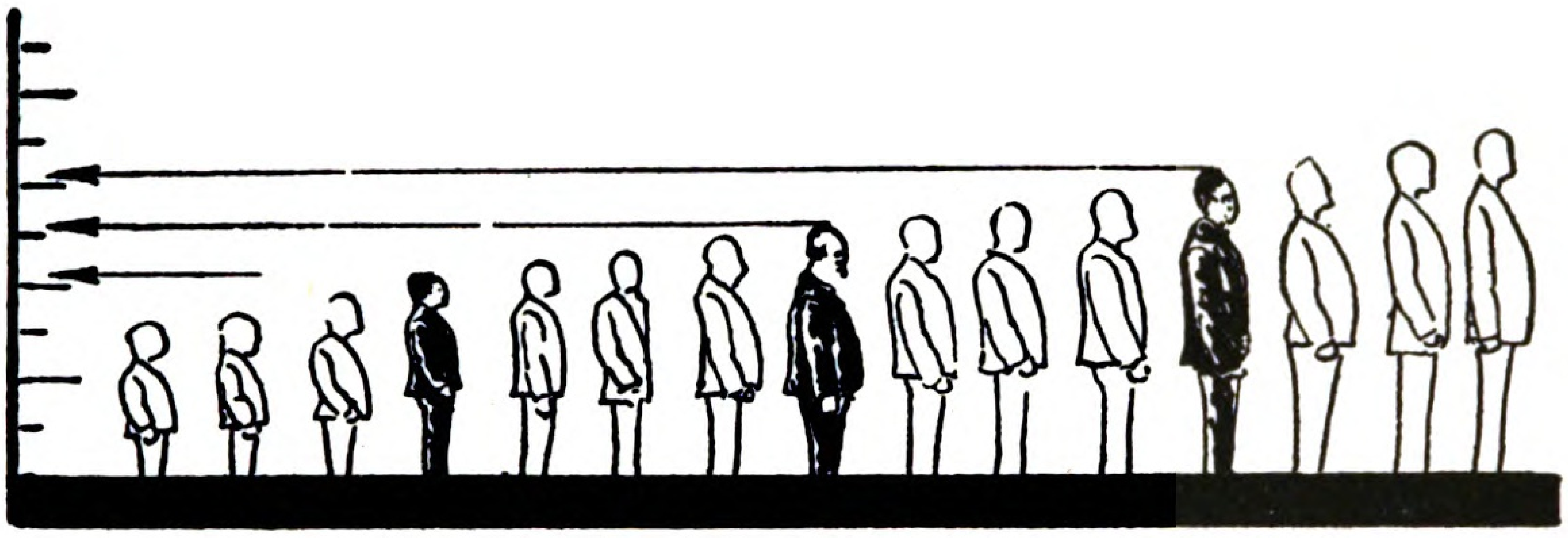
The Alpha Corporation may use all three averages to give a fair picture of the wage situation.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Every one of them has its particular characteristics and areas of usefulness.

Instead of describing the height data by one typical expression, we may use several typical values to characterize the height data of the Alpha workers. We may, for instance, mention the height of three men.



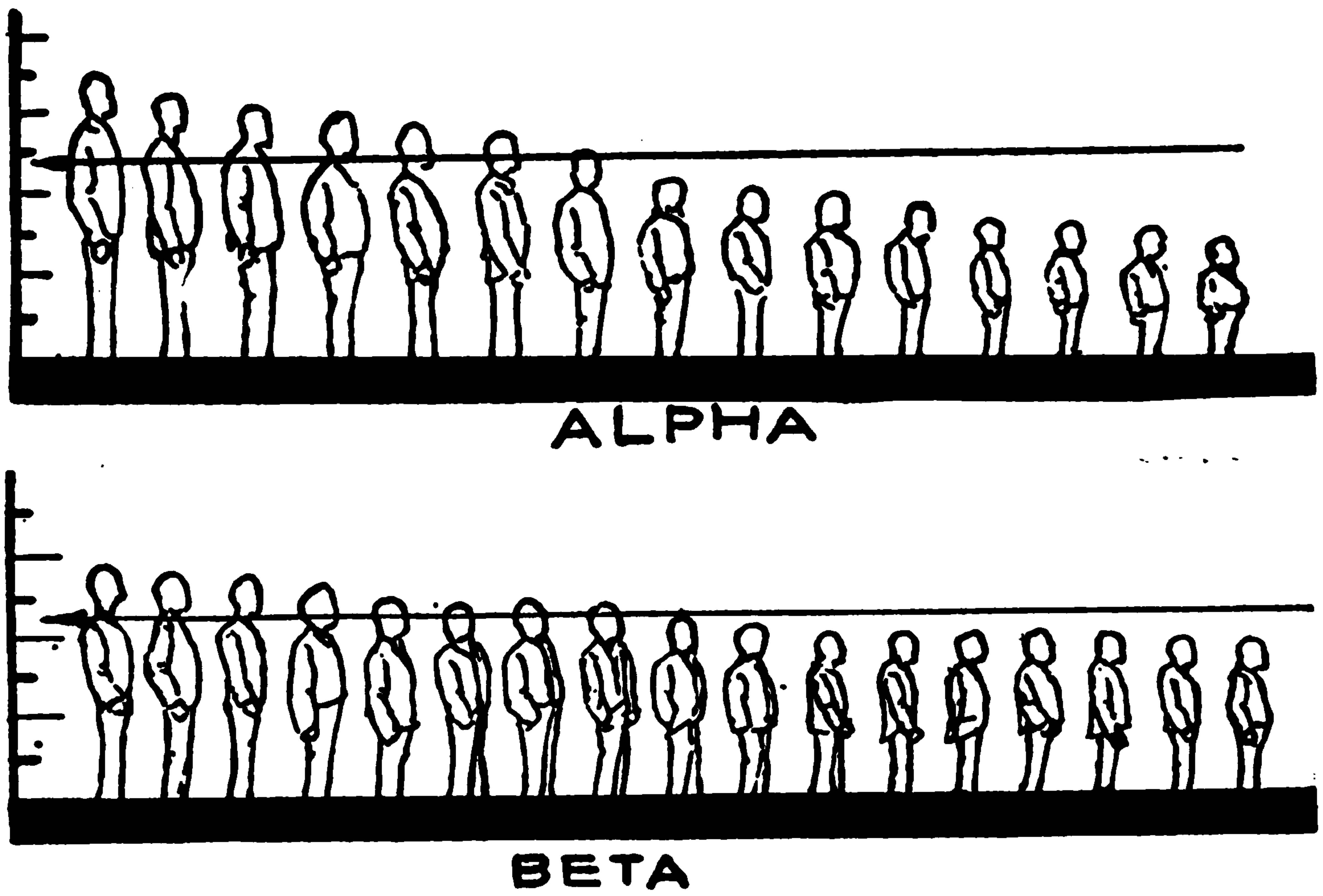
The first one—his height is called the first quartile—is equalled or exceeded in height by ¾ of all the men. The second is known to us; he is the middleman and his height is the median height. The third one, whose height is the third quartile, is equalled or exceeded in height by ¼ of all the workers.

The Second Shorthand Expression

We are describing the height situation of the Alpha workers by, let us say, using the arithmetic

Digitized by Google Original from UNIVERSITY OF MICHIGAN

mean. It so happens that we also know the height situation of male workers in the Beta Corporation. The mean height is the same as in our Alpha case. If we were to describe the height situation simply by the first shorthand expression, we would create the impression that the height situations in both corporations were alike. A closer investigation, however, shows a considerable difference.



The Alpha Corporation has workers ranging from tall to short, whereas the Beta Corporation has only workers of approximately "average" height.

Our first shorthand expression has given us a good, but incomplete description. We shall have to

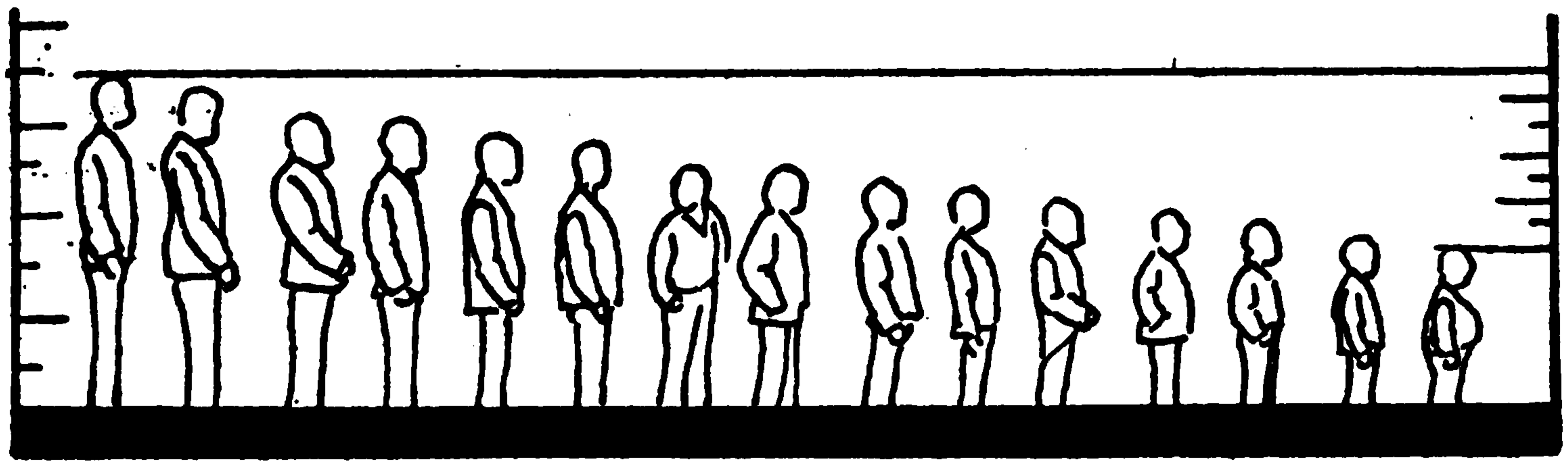
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

describe the differences in height from the average height, or the so-called variation about the average, which is our second shorthand expression.

This expression will show us that there is a small variation about the average, or, as we may also say, a considerable uniformity in the case of the Beta Corporation. That makes the average a really typical representative of the whole group of workers.

In the case of the Alpha Corporation, with tall as well as short men, or a considerable variation about the average, the average gives a less reliable description of the whole situation.

If we wish to indicate the variation or dispersion in the height of Alpha workers, we may state the difference between the tallest and the shortest man.



This difference is called the range. It is not the difference to any average, but it can be used to give a rough idea of the variation.

In the case of the Beta Corporation, the range is, of course, much smaller.

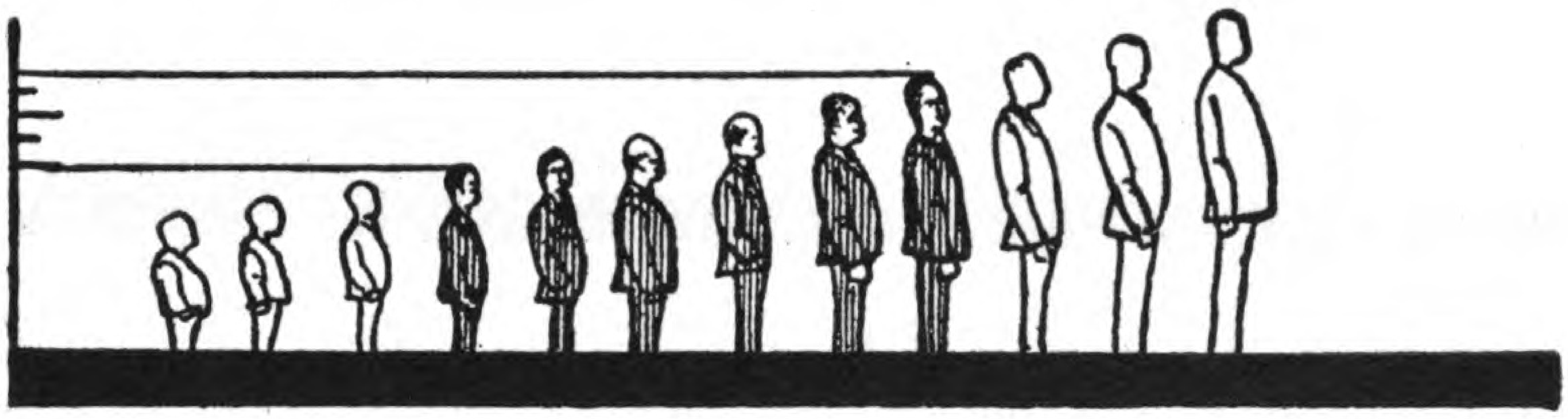
Digitized by Google Original from UNIVERSITY OF MICHIGAN

If we had just one very tall or one very short worker, the range would be different. This shows us that the range is not a reliable measure as just one out of even a great number of workers may change it considerably.

The range is frequently used as a measure of stock market fluctuations.

The idea of the range will be somewhat improved if we disregard both ends of our line of workers.

We may, for instance, exclude ¼ of the men at the tall end and ¼ at the short end. This most probably will take care of the atypically tall or short workers.

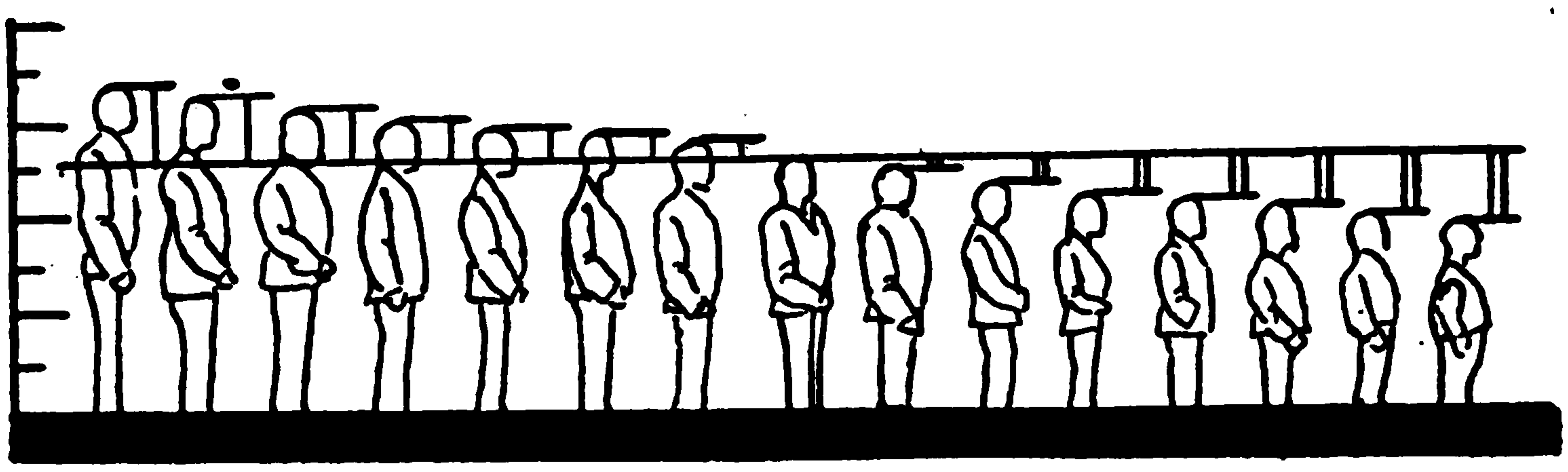


In other words, we are taking the range between the quartiles, and not between the tallest and the shortest man. This is called the interquartile range. Customarily, ½ of this range is used as a measure of variation and is called the semi-interquartile range or the quartile deviation.

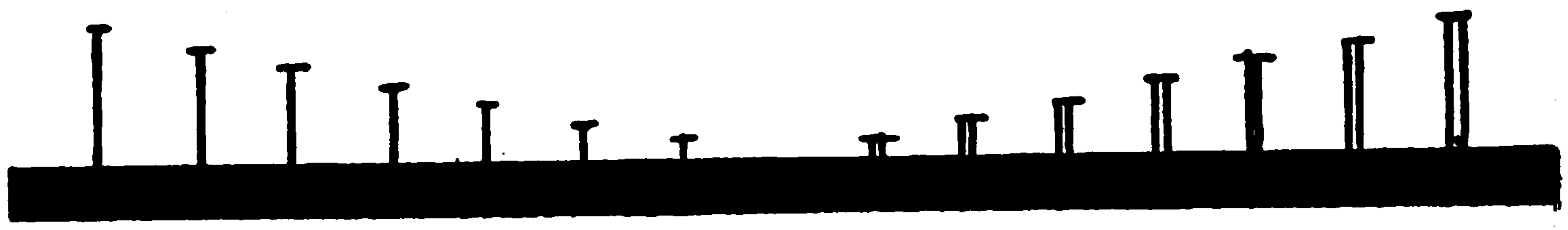
Instead of indicating the variation which exists

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

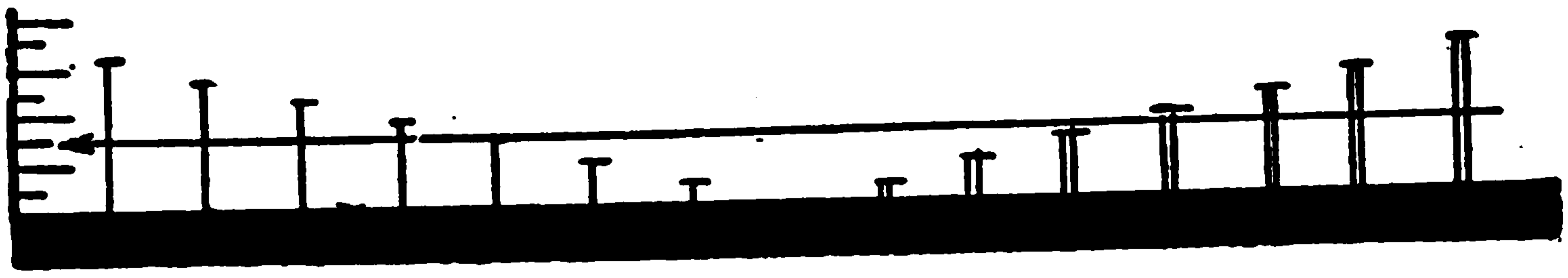
in the height of the Alpha workers or of the Beta workers by mentioning some type of a range, we may consider how much every single worker varies from the average height.



Let us look at all these individual differences from the mean. By the way, we can also study the differences from the median, if we wish. Here are all these height differences again:



Let us now take the arithmetic mean of these differences.

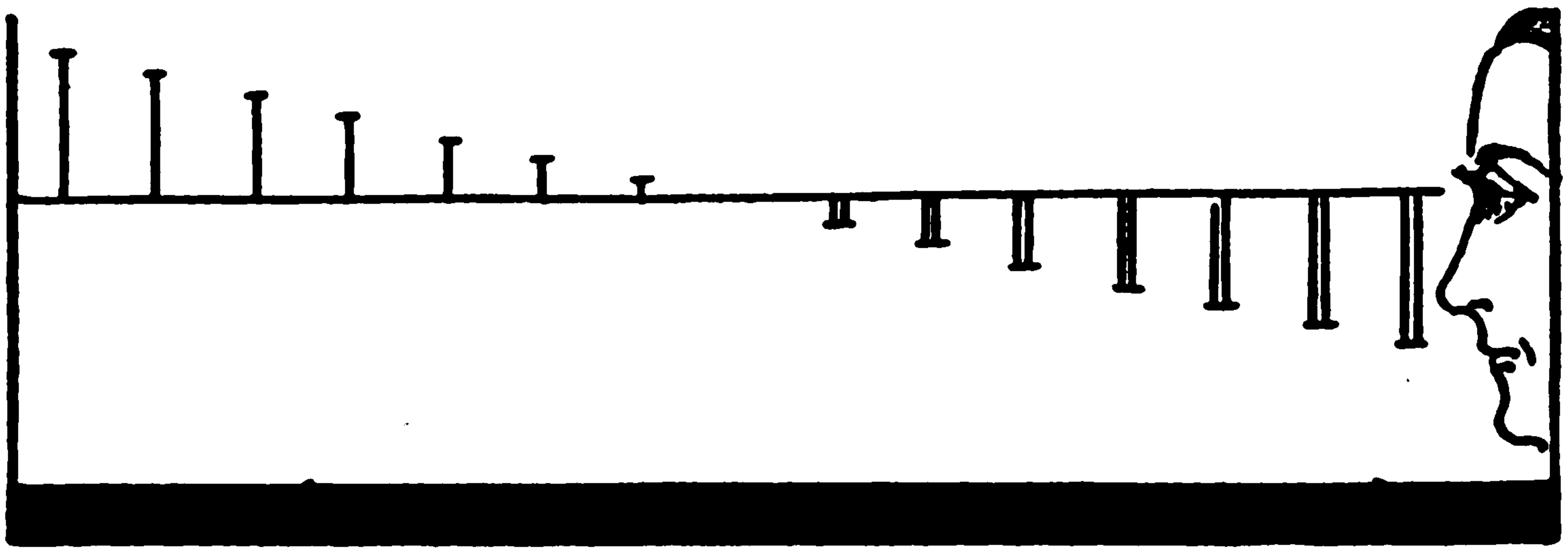


It shows us how much a worker varies on the average from the average height. This mean difference is called the mean deviation or average deviation.

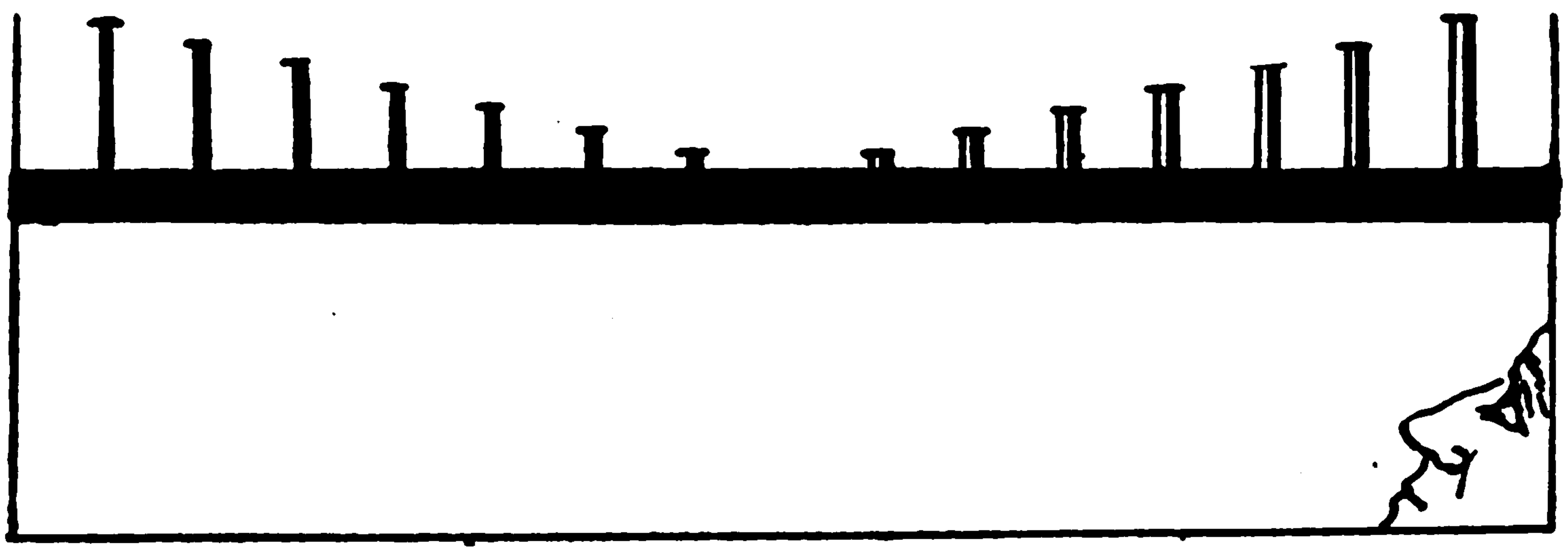
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

We can see readily that the mean deviation for the Beta workers would be smaller than for the Alpha workers.

There is an objection to this measure of variation. We have been treating the height differences in the direction of taller-than-average and shorter-than-average alike. Many statisticians think we should distinguish, in our reasoning, between the differences above and below the average height.



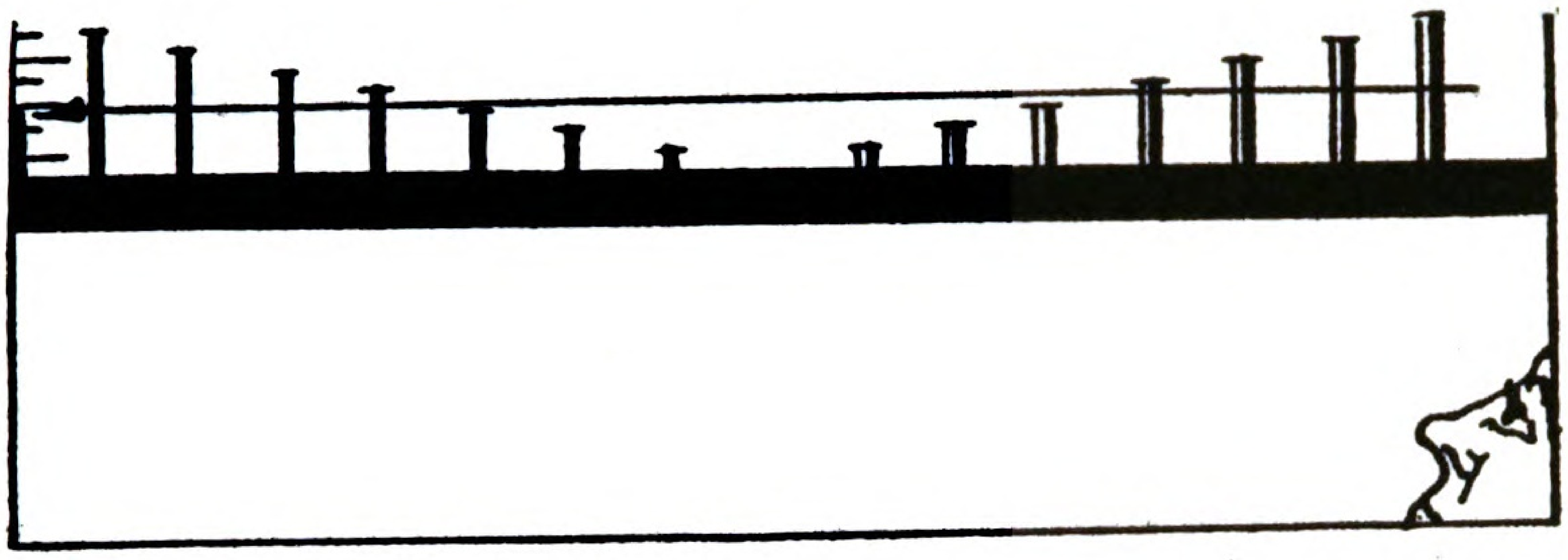
The distinction between above and below differences, however, will cease to be a problem if we put all differences upon a higher level.



Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

We may now, as before on the lower level, take the arithmetic mean of the differences, and this



mean is called the variance. The variance, brought back to the original level, is the standard deviation. It is not only the most valuable measure of variation but also one of the most important statistical concepts.

The standard deviation, as a good measure of variation, will give us an accurate idea of the scatter or dispersion of the height data about the mean height. A large standard deviation for the Alpha Corporation will indicate that the height distribution of the workers is less uniform than the distribution of the Beta workers.

If we were just to see the standard deviations for the height of the workers of these two corporations without seeing all height data, we would know that the Alpha Corporation has workers ranging from tall to short, whereas the Beta workers are all close to the average height.

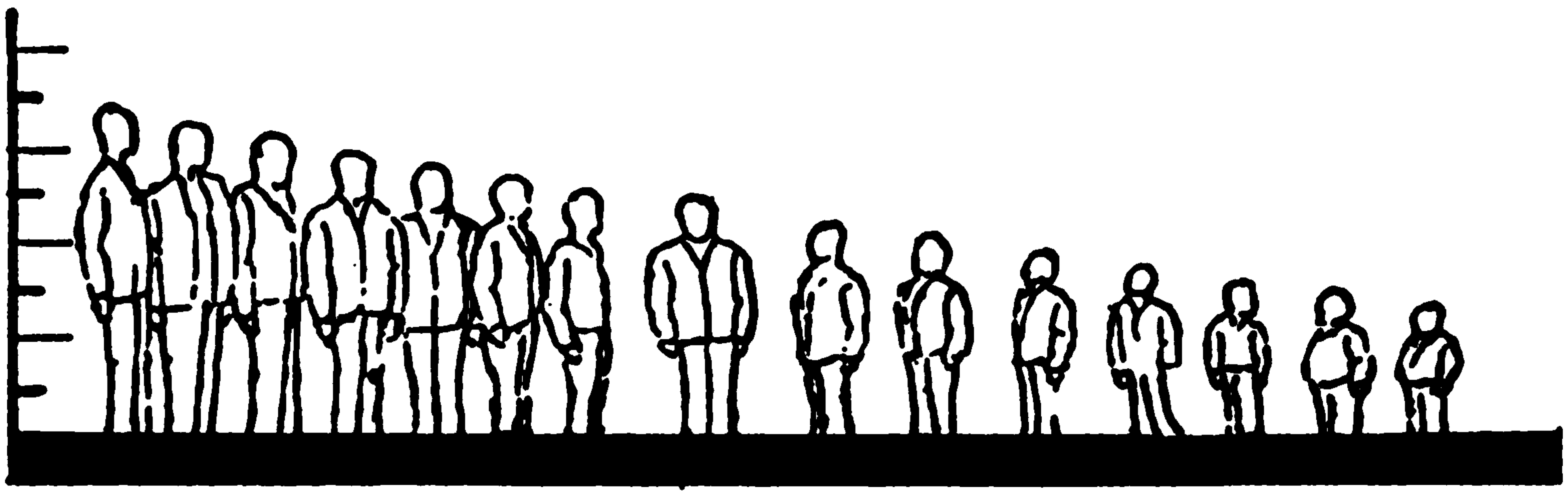
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Or let us look at wages. The wage distribution of all Alpha workers shows a greater average deviation as well as standard deviation than that of all Gamma workers; the average wage paid in both corporations is about the same. That means that there is a comparatively greater variability in Alpha wages. Wages paid by the Alpha Corporation are scattered widely about the average wage.

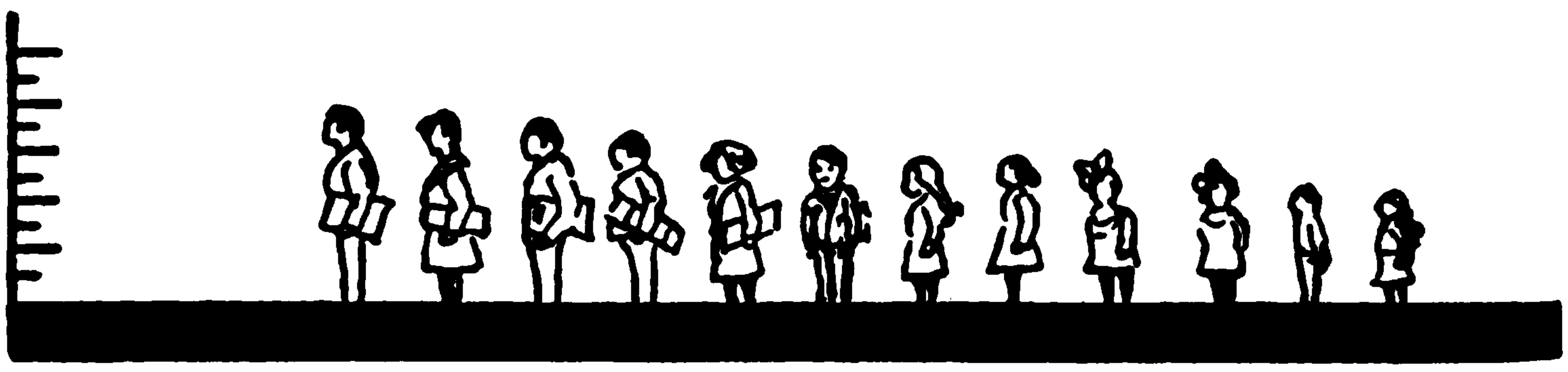
Gamma wages, on the other hand, are more uniform; no worker receives much more or much less than the average wage. Obviously, the average wage of the Gamma Corporation would represent the wages at Gamma much better than the average wage of the Alpha Corporation would represent the Alpha wages.

There are situations, however, where the standard deviation cannot be used in its original form. If we were interested, for instance, in comparing the height situation in the Alpha Corporation and the height distribution of 6-year old school children,



Digitized by Google

Original from
UNIVERSITY OF MICHIGAN



we might readily see that the two standard deviations are not comparable. They are based upon the distances to entirely different means. Such distances of course have a different significance.

A second situation where the standard deviation alone would not help us is present if we, for instance, want to compare the height and weight situation of the Alpha workers. We want to know whether there is greater uniformity in height or weight. Standard deviations expressed in inches and in pounds obviously cannot be compared.

In situations such as the two mentioned, the standard deviation can be used to compare variation—or uniformity—if we express the standard deviation in terms of the respective mean; thus we may say, the standard deviation is a certain percentage of its mean. Now it does not matter any more that the means are different with regard to magnitude or the units in which they are expressed. The standard deviation of the school children's height is expressed in terms of the mean height of the school

Digitized by Google

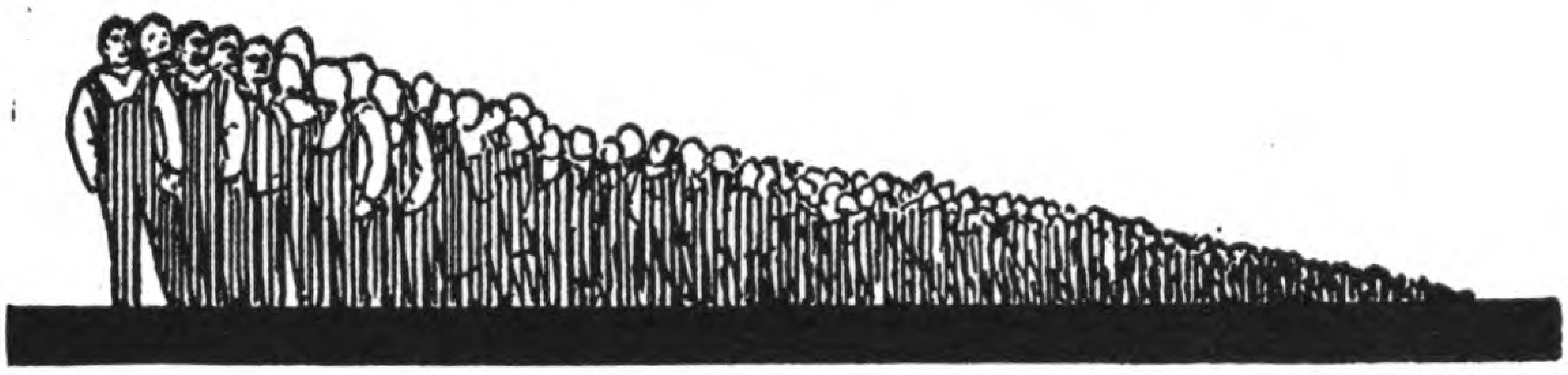
Original from
UNIVERSITY OF MICHIGAN

children, the standard deviation of the weight distribution in terms of the mean weight, and so on. Thus we are in our comparisons independent of the actual features of the means, their differences in magnitude or units.

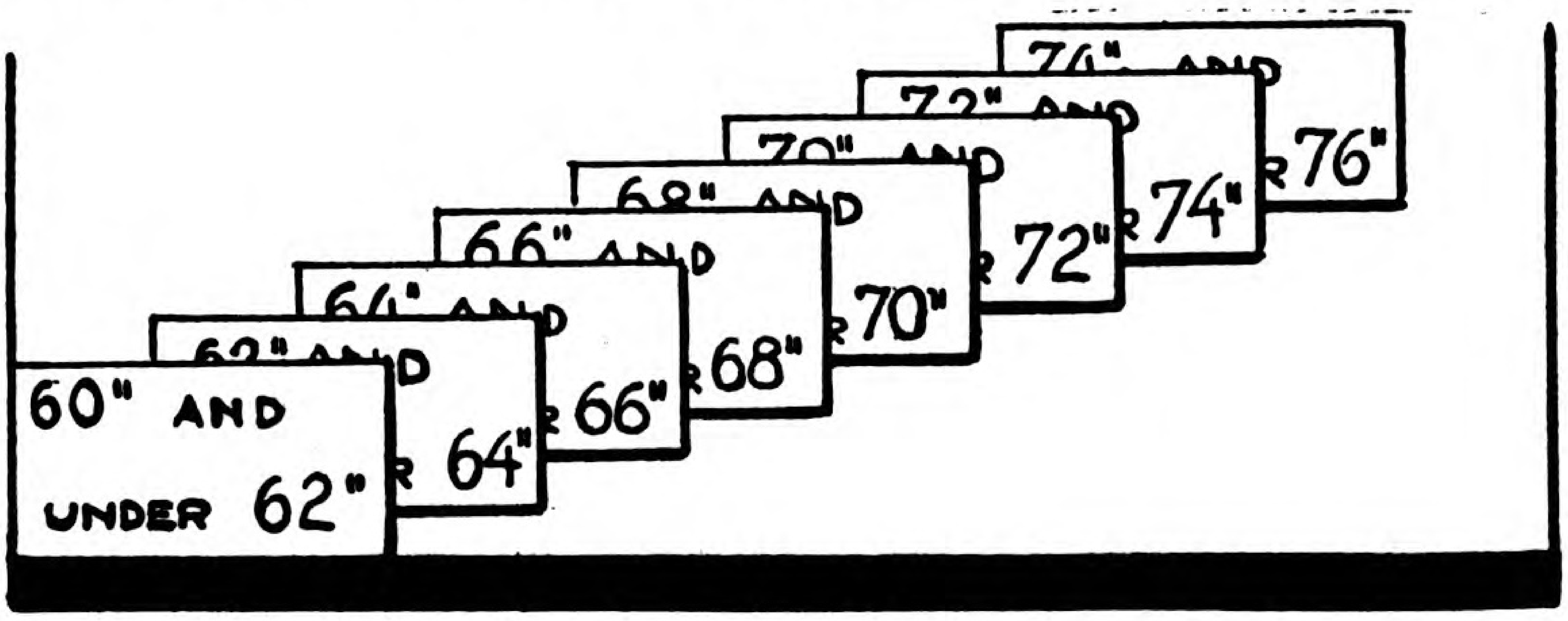
The standard deviation, transformed in the way just described, is called the coefficient of variation.

More About the First Two Shorthand Expressions.

The Gamma Corporation has a very large num-

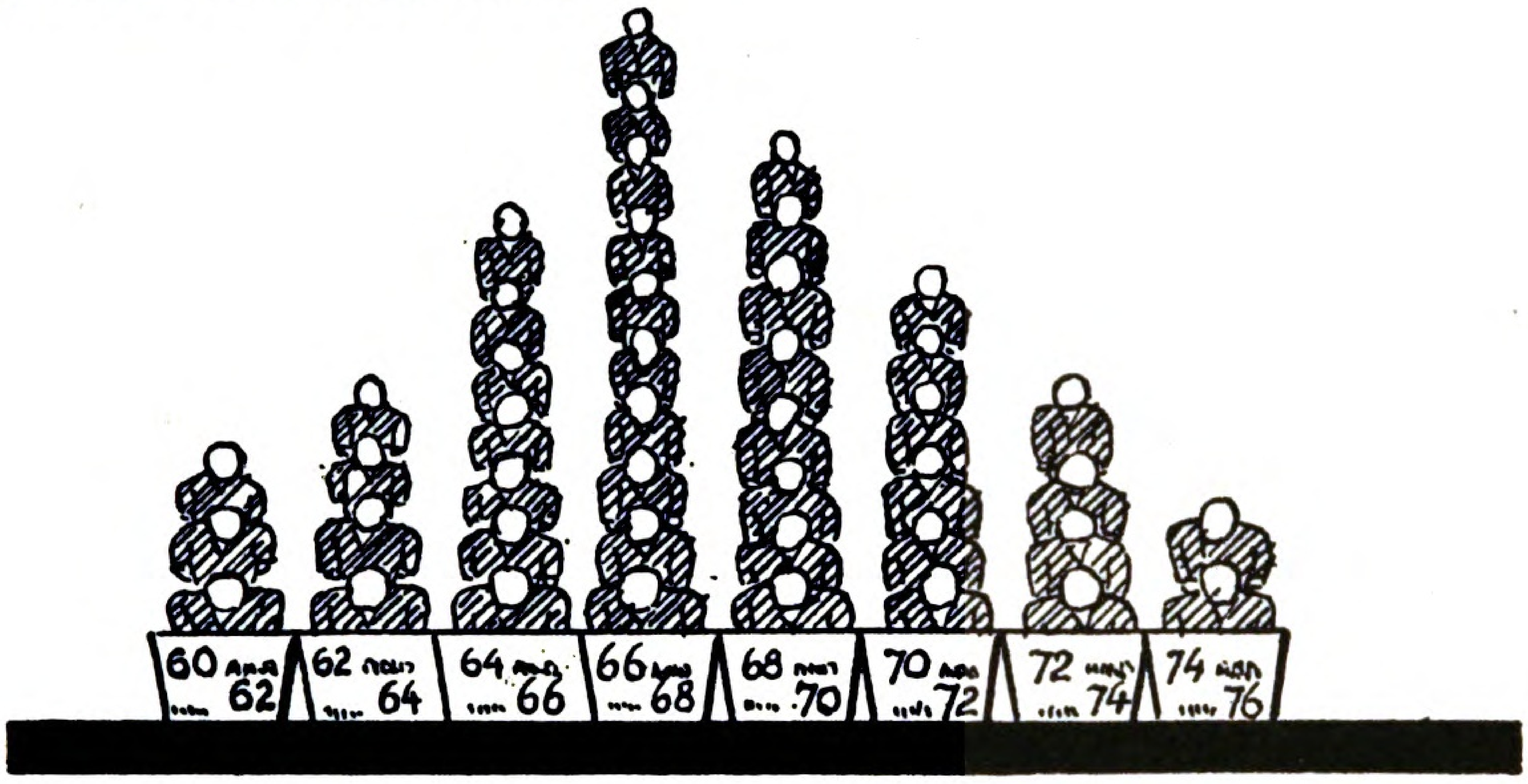


ber of workers. It would be quite laborious to deal with such a long line of men. We simplify the situation if we set up a number of signs and ask the

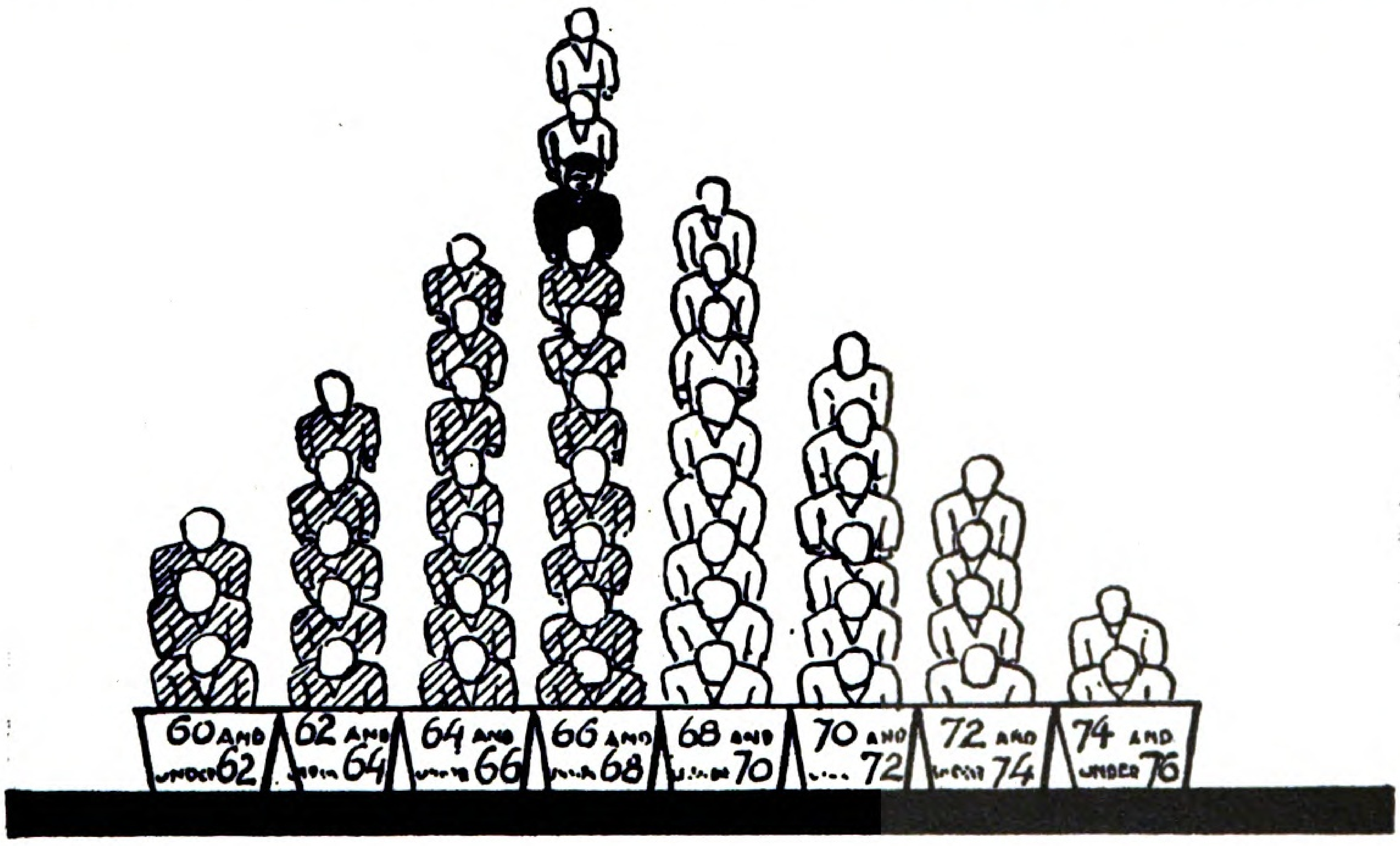


Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

workers to line up behind the appropriate sign. From a window on the 10th floor the mass of male workers would look like this.



This arrangement of people or items into groups is called a frequency distribution. The most common height is that of the men in the 4th column from the left. The greatest number of workers has



Digitized by Google

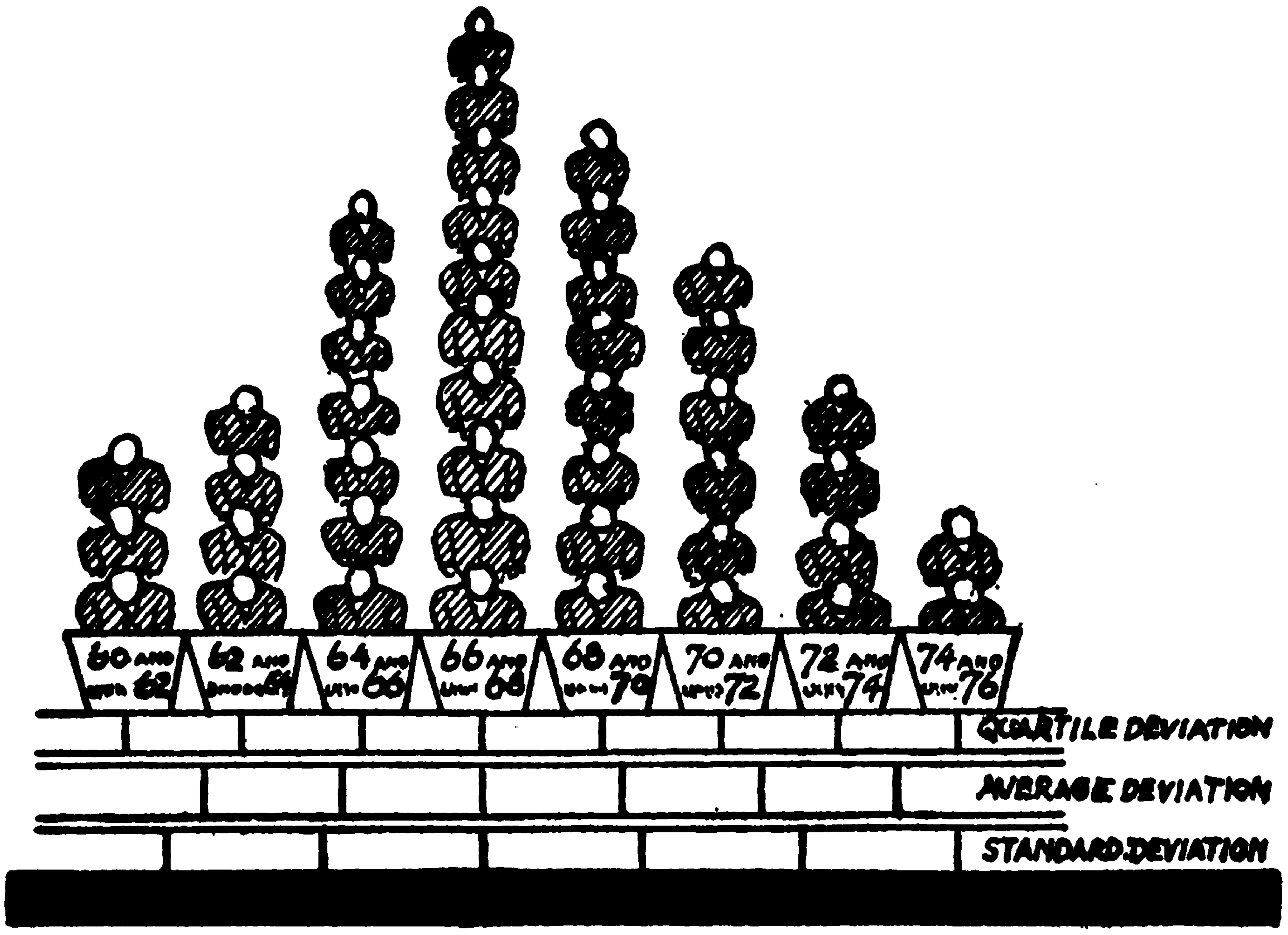
Original from
UNIVERSITY OF MICHIGAN

this height and therefore the mode will be somewhere between the limits of this group.

The height of the middleman, the median, and the mean height which balances all the deviations from it may be found.

The variation may be indicated by the range, the distance between the lower limit on the leftmost and the upper limit on the rightmost sign. The quartile deviation again is based upon the distance of the quartiles, and the average deviation upon the average distance between a worker and the mean or median.

We may think of averages as points whereas all measures of variation represent a distance.

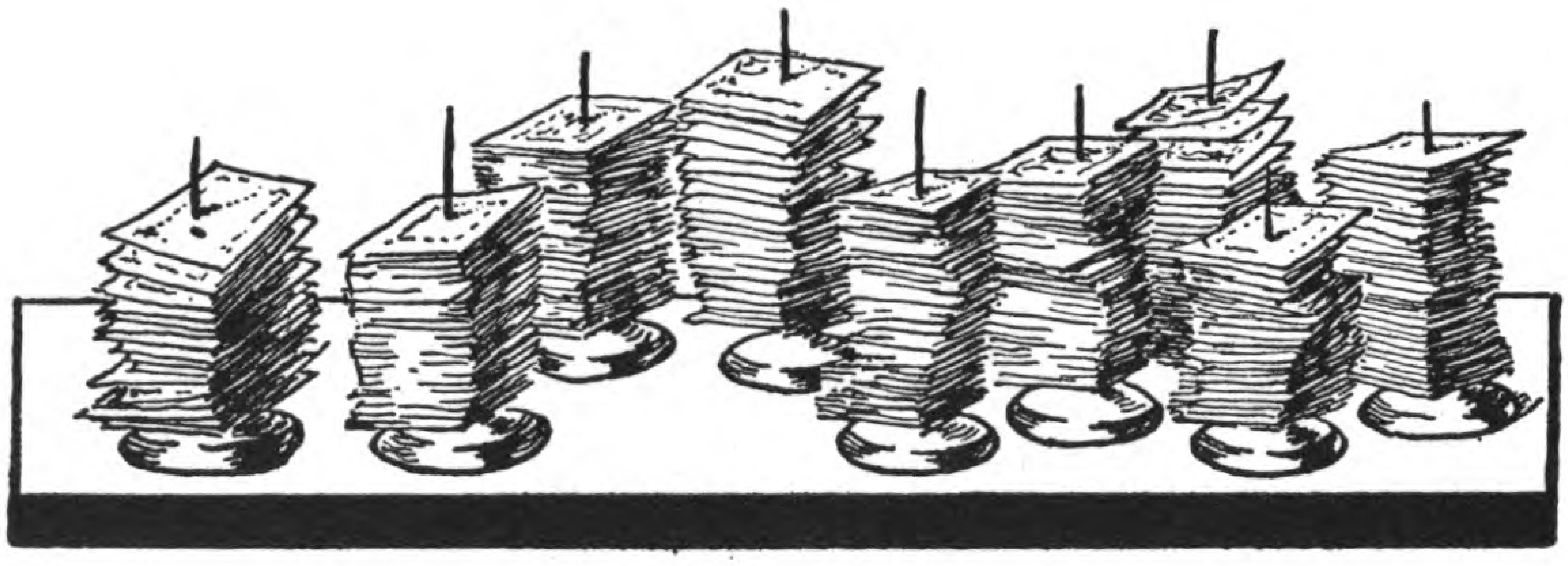


Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Since all of these measures use different approaches, a quartile deviation is comparable only to another quartile deviation and a standard deviation only to another standard deviation.

Up to now, we have been interested in the height situation of workers. What we saw about the statistical aspects applies, of course, to a very large number of other problems. We may be interested in wage data or prices, age data or IQ levels, income data or sales.

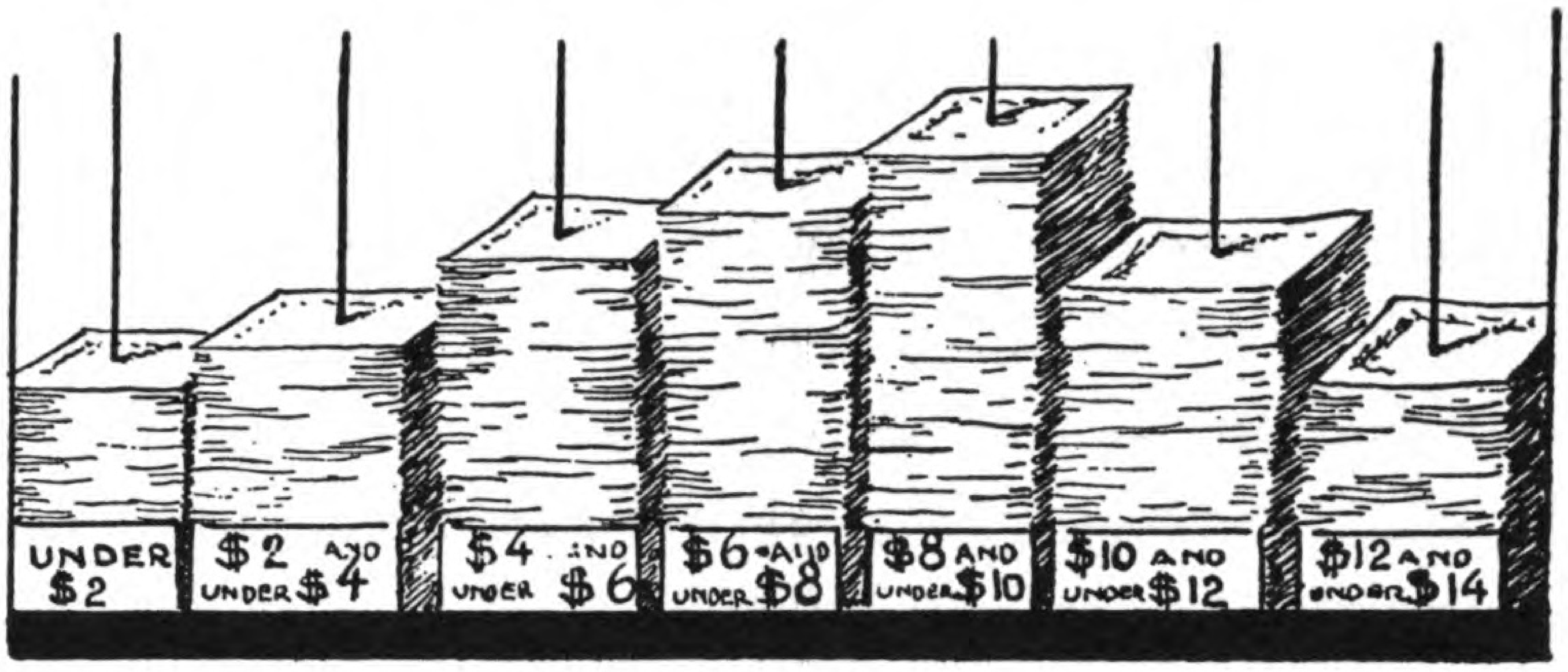
Let us take sales figures as an illustration. At the close of a busy day the sales checks of Alpha's retail stores look like this:



We may sort the sales checks by putting all checks under \$2 in one batch, the sales checks between \$2 and just under \$4 in the next, and so on. Again, any

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

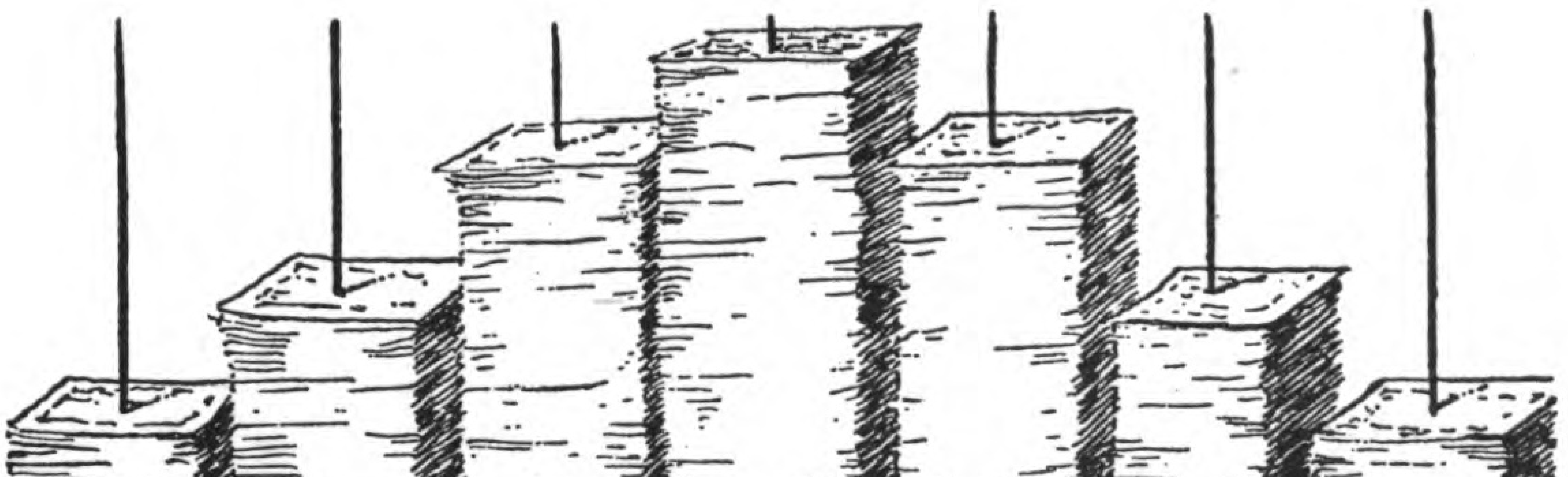


ALPHA

measure that we wish to use for the first or the second shorthand expression may be ascertained. However, we are interested now in the third and the fourth shorthand expressions.

The Third and the Fourth Shorthand Expressions

Let us compare one day's sales checks of the Alpha Corporation and of the Beta Corporation.



BETA

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

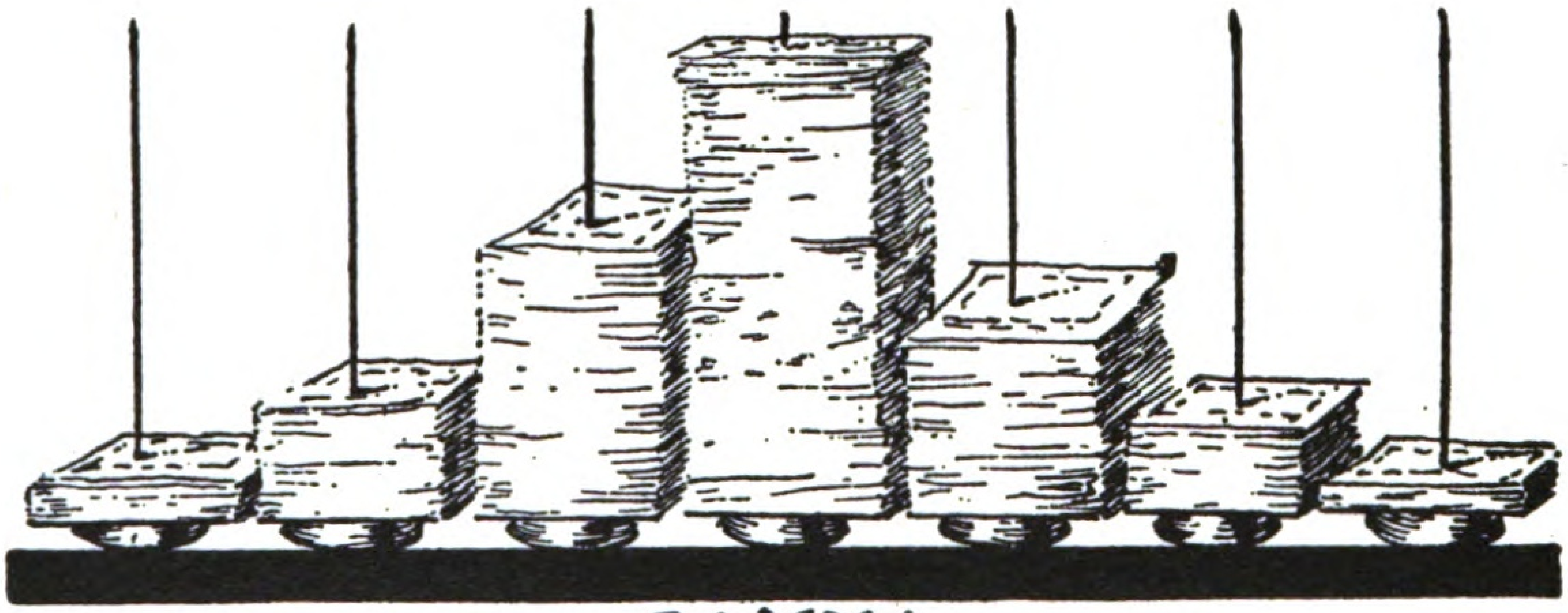
The arrangement of the Beta sales checks is symmetrical, the one of the Alpha checks is not. Any distribution that is not symmetrical is called skewed. The sales checks of the Alpha Corporation show a distribution that is skewed to the left since the batches of sales checks tail off to the left. There are fewer checks of the lower amounts.

Skewness is a characteristic trait of a distribution and measures of the degree of skewness constitute our third shorthand expression.

If we wish to characterize a distribution, the degree of skewness is an essential part of the description.

The sales checks of the Beta Corporation, of course, show no skewness.

The sales checks of the Gamma Corporation differ from the distribution of the Alpha checks on another point.



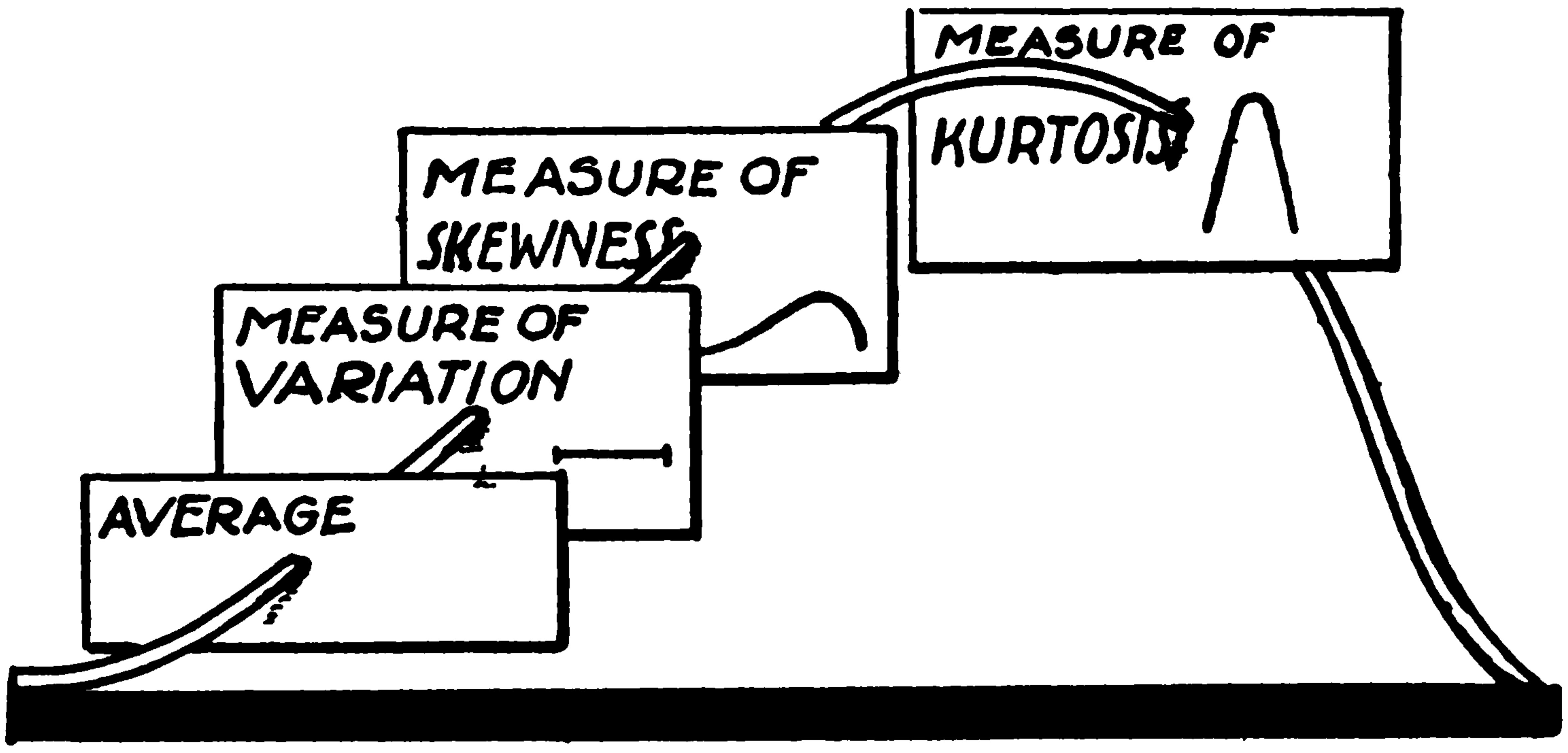
GAMMA

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The average sale may be the same in both corporations, but the Gamma sales checks certainly bulge to a greater extent. Statisticians use the Greek word for bulging, kurtosis, to designate this fourth characteristic of a distribution. Our fourth shorthand expression, therefore, shows us the degree of peakedness that a distribution exhibits.

We are now in a position to describe adequately any group of data which differ among themselves by characteristics that can be counted or measured. Inches, pounds, etc., can be measured; dollars, years of age, etc., can be counted.

We can describe any frequency distribution for the purpose of comparisons, by using our four shorthand expressions.

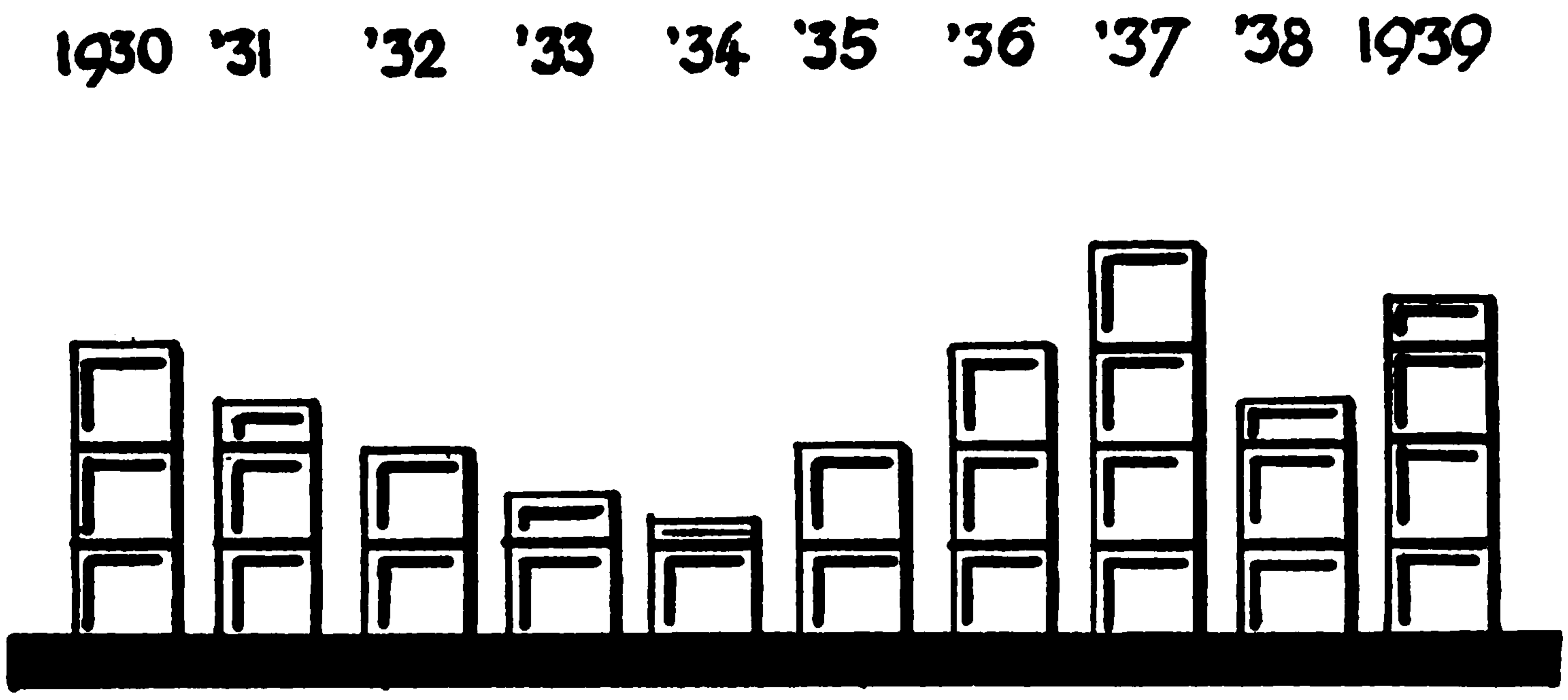


Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

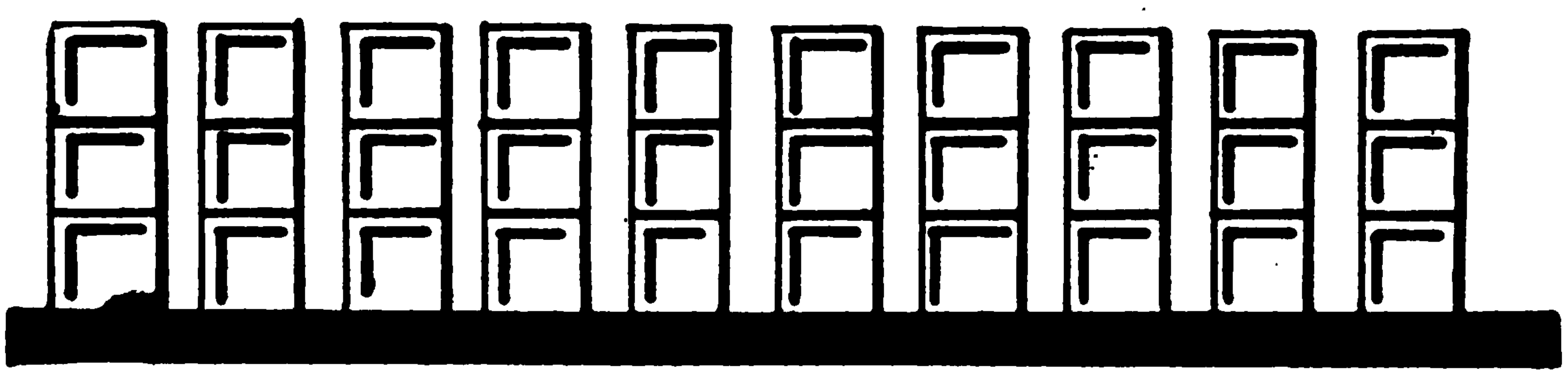
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Time Series

The Alpha Corporation shows the following production during the 1930's:



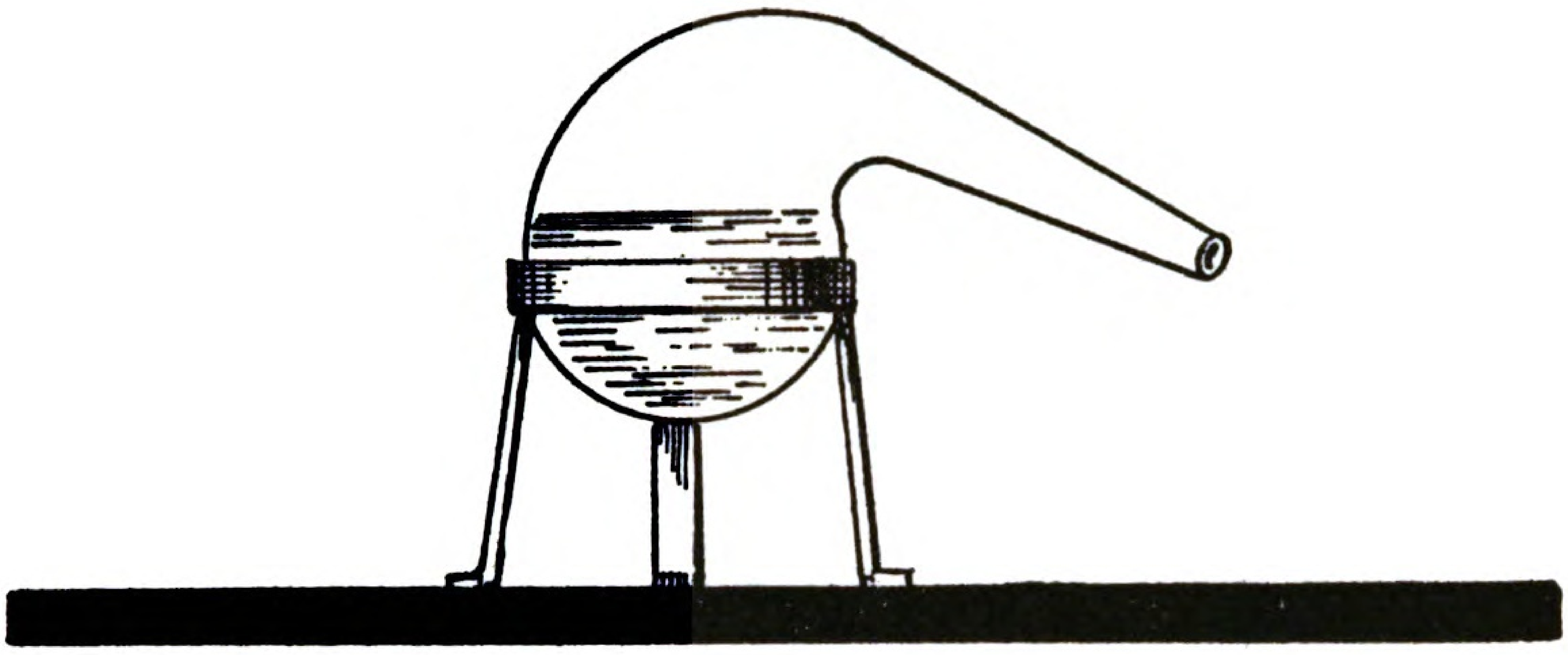
Why does it not look like this?



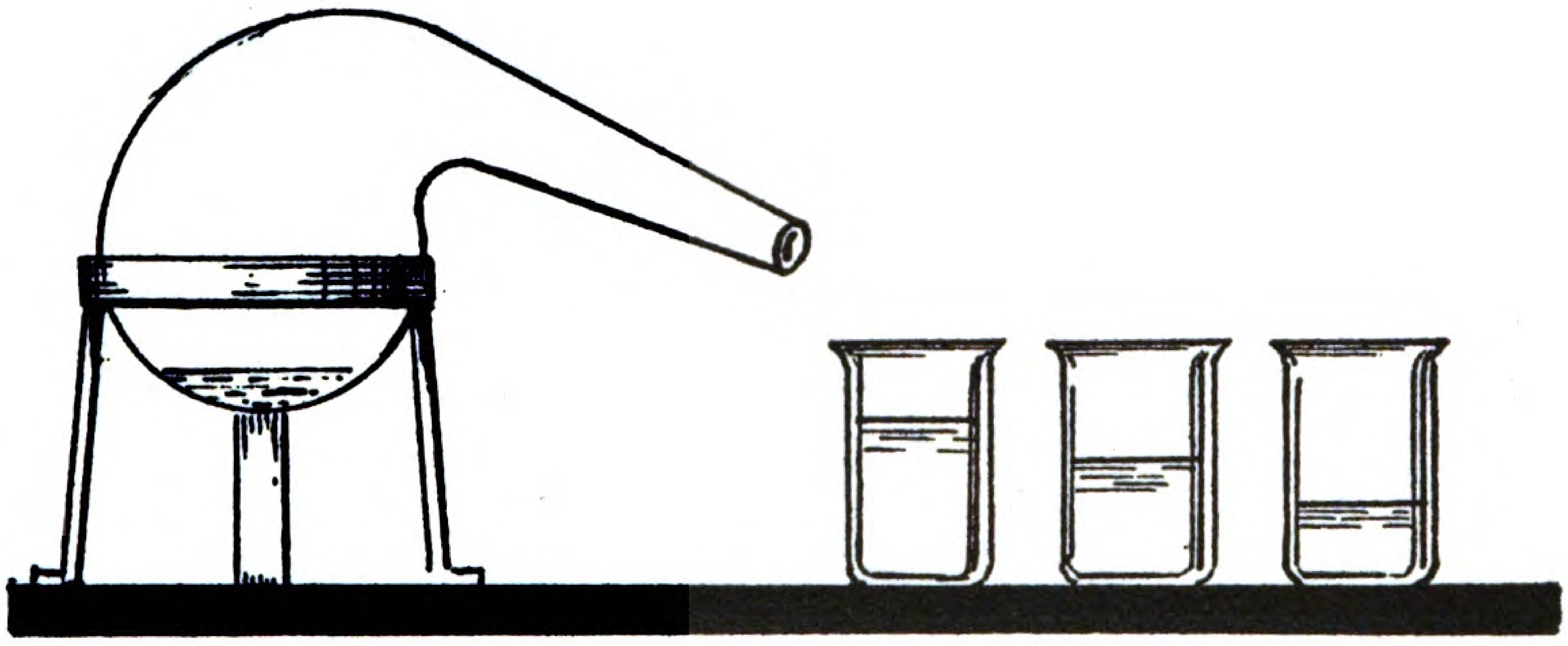
There is obviously a force that pulls it up or pushes it down and finally brings about the production picture that we actually find.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Let us try to analyze this force. We may use the methods of a chemist who tries to analyze a liquid.



He may first separate one element, then a second and a third. In his test tube he may find a residue that defies further analysis.



The force that influences the production of the Alpha Corporation is a blend of four elements,

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

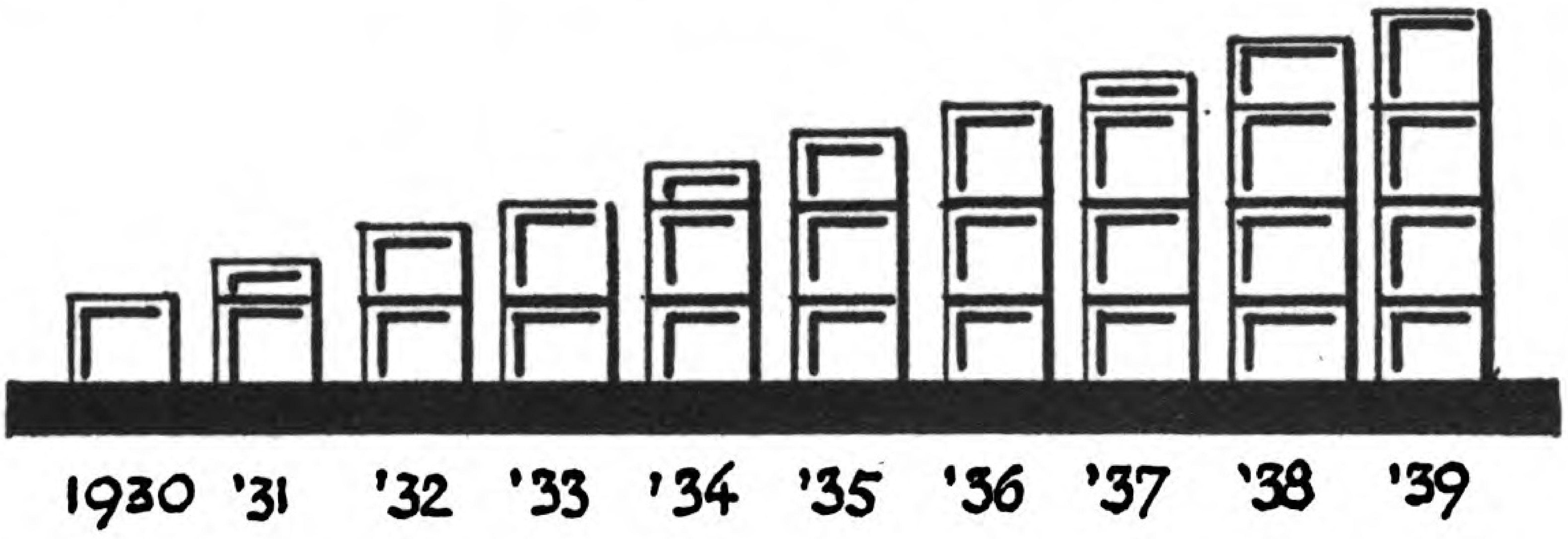
one of which is a residual factor that cannot be analyzed.

Let us now start to take this composite economic force apart; let us isolate its basic factors.

The First Element

The production of the Alpha Corporation has a tendency to expand steadily over a period of time. The population grows, which means there are more and more customers; technological changes take place and business organization in general improves. These are just a few reasons for the increased production of the Alpha Corporation; there are more.

If this tendency to grow were the only factor that influenced the production of Alpha, the production may have shown the following development:



This growth factor, however, which we call the trend, is just one of four elements.

Production of the Beta Corporation shows a

Digitized by Google
Original from UNIVERSITY OF MICHIGAN

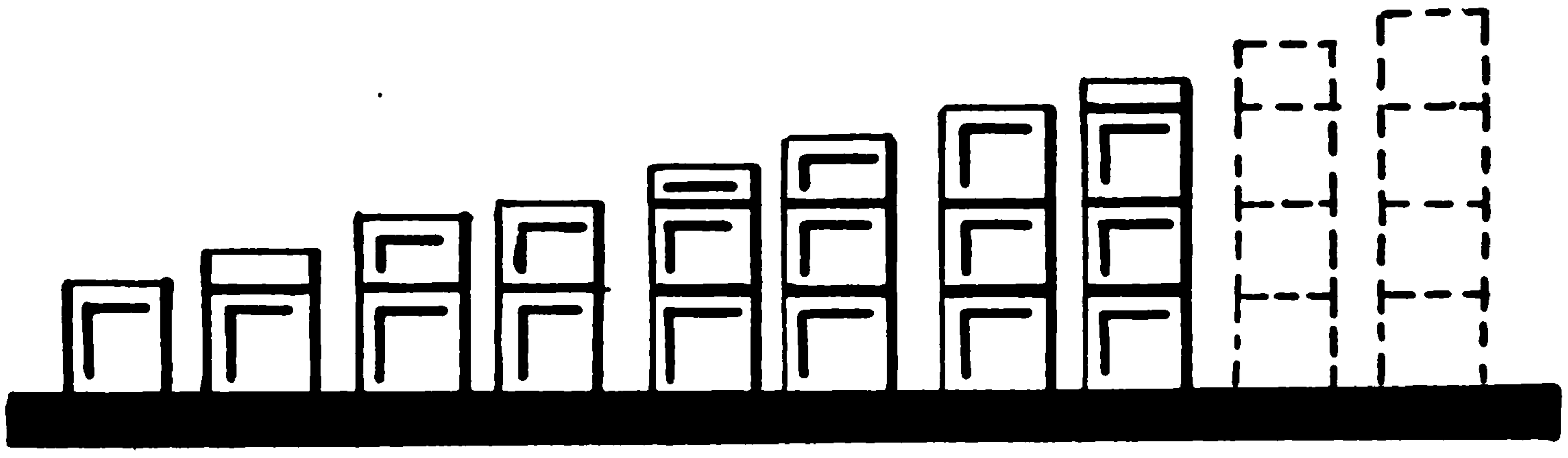
steady tendency to decrease, because of changes in demand. This negative growth factor is also called a trend.

Increase or decrease does not have to be consistent. The growth factor may change over a period of time. The production of the Gamma Corporation for instance showed a tendency to grow rapidly for a number of years, then just slightly, and later showed neither a tendency to increase nor to decrease. Recently, production showed a tendency to decline. This growth factor in changing direction is also a trend. We call this type a curvilinear trend.

We may wish to separate the trend from the other elements that influence the Alpha production over a period of time, because we may want to analyze the remainder, isolating one element after the other, or we may want to study this first element itself.

We may be interested in knowing and studying the growth factor of the Alpha Corporation, we may wish to compare this growth factor with others, for instance the one for the entire industry, or we may attempt to extend this growth factor into the future. This is called forecasting and presupposes that we have clearly recognized the tendency to grow and that this tendency will continue unchanged.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN



THE SECOND ELEMENT

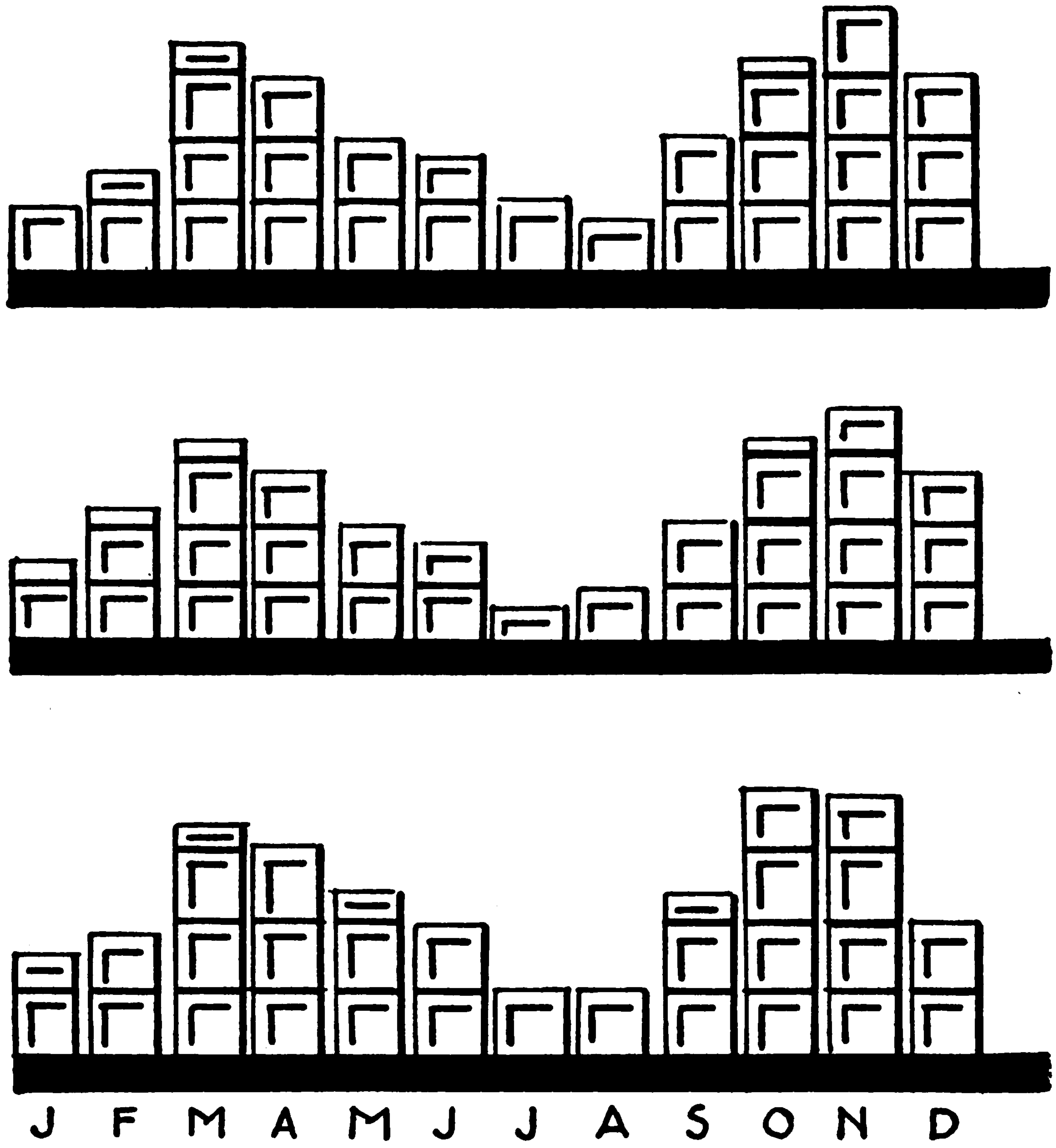
We have seen the annual production of the Alpha Corporation. It may be worth while to look into the month-to-month development. Let us take any period of three years after trend has been eliminated. The monthly production may shape up as shown on the following page.

The monthly production of the Alpha Corporation varies considerably, yet there is a pattern which we find repeated in each year and which generally is due to the influence of climate and customs. We designate these fluctuations as seasonal variations.

The large sale of Alpha products before Christmas accounts for the increased production in October and November. The hot summer weather causes a reduction of sales and of production.

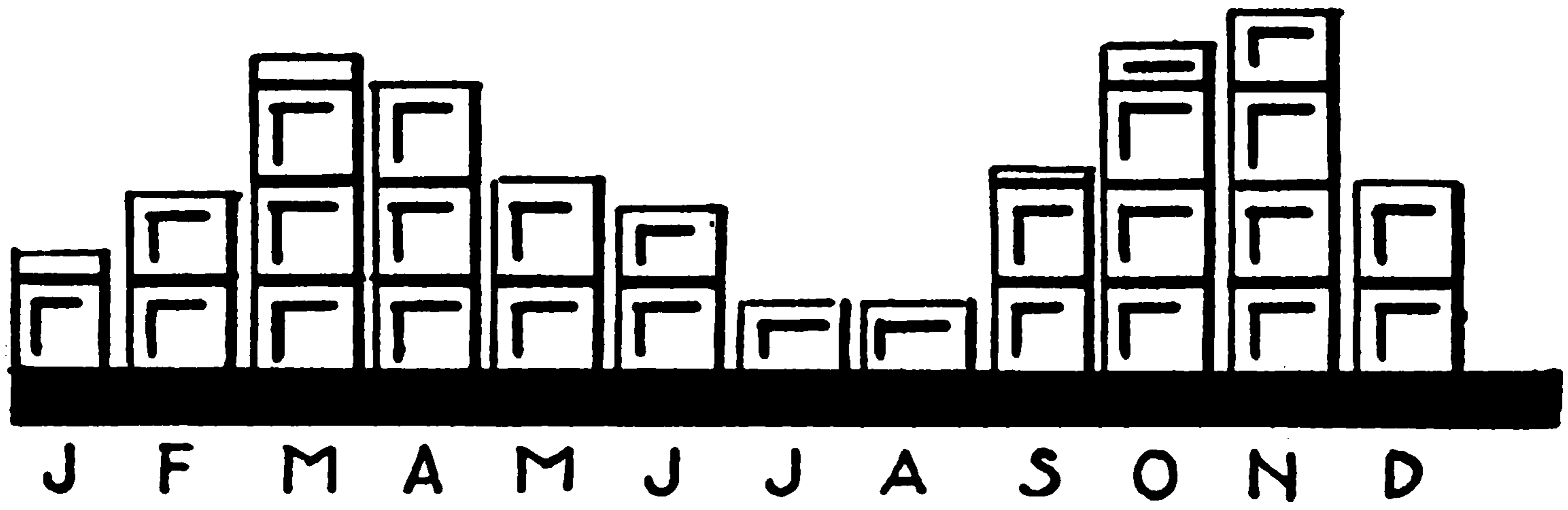
The situation has not been exactly the same in each year. An inclement winter, early Easter, etc., may have influenced the intensity of seasonal variations.

Digitized by Google Original from UNIVERSITY OF MICHIGAN



Since we are interested in developments over a period of time, we shall, however, consider the seasonal position of the average January, the average February, etc.

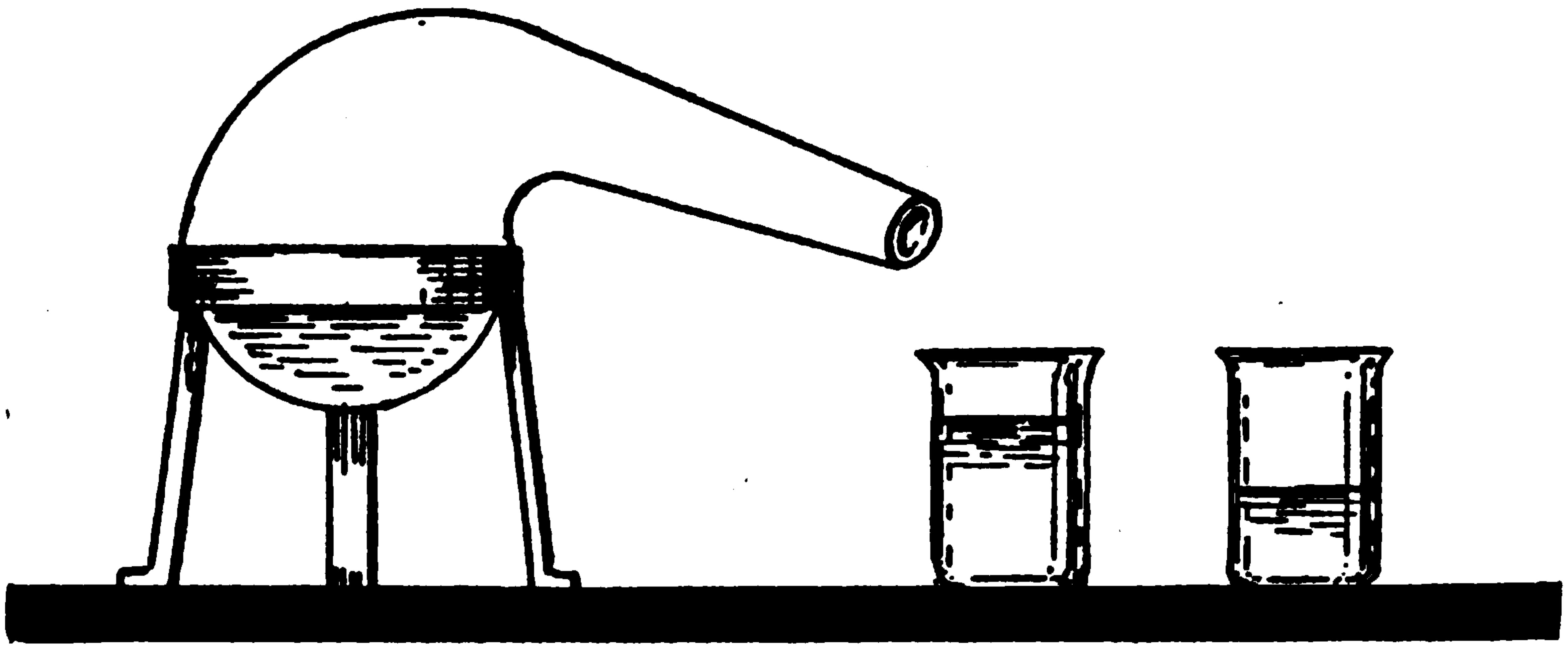
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN



This typical seasonal development is called the seasonal pattern. The actual production of the Alpha Corporation will tend to follow the pattern.

We may find that such a pattern changes gradually or abruptly. The Alpha Corporation may take up the manufacture of a new item which is produced mainly in summer. That would change the pattern abruptly.

Again, we have two reasons for our analysis of



seasonal variations, to separate it from the other elements or to study it. That means, we may isolate

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

this second element as a necessary step in the process of decomposition, in order to find the other elements; another purpose would be the study of seasonal variations themselves. The management of the Alpha Corporation will have to take into account seasonal variations in many decisions. Labor will suggest that these fluctuations be ironed out, in the interest of steady employment. Any forecast of production for next May or October has to consider the seasonal position of May or October.

The Third Element

The economy of the entire country experiences changes; there are "good" times and "bad" times. Economists talk of a wavelike up and down from prosperity to crisis, depression, recovery, and on again to prosperity.

The Alpha Corporation follows a similar cyclical development. The waves are generally less pronounced as compared with the business cycle of the entire country and the rhythm of the waves shows a time lag of a few months in the case of the Alpha Corporation, but generally the cyclical development shows a family resemblance to that of the entire country and to that of other similar corporations.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The power that increases the Alpha production in times of prosperity and reduces it in times of depression constitutes the third element in the force that influences the production of the Alpha Corporation. This power caused the 1938 production to be smaller than the production of the preceding and the following years.

It is rather difficult to analyze "the cycle," since one cycle differs widely from the other. Nevertheless, something like a cyclical pattern of a country or an industry can be established and can be utilized for forecasting. Of course very great care must be exercised in using such cyclical patterns.

Usually we eliminate the first and the second elements and leave the third combined with the fourth. We then try to separate mechanically the third and fourth elements as much as we can.

The Fourth Element

The growth factor, the seasonal and the cyclical fluctuations together do not account for all variations in the production of the Alpha Corporation. There was a strike that reduced production during one month; fire and flood influenced production in other months. Such events, which occur without

Digitized by Google Original from UNIVERSITY OF MICHIGAN

any regularity, cause "irregular fluctuations." They constitute the fourth element.

Irregular fluctuations show no pattern, not even a family resemblance. We cannot isolate them, as we can the other three elements, but we obtain the irregular fluctuations as a residue. Not the entire irregular factor, however, remains as a residue; many small parts are attached to the other three elements and cannot be removed by the tools that are at our disposal. They remain as impurities.

Combinations of Elements

We have now the four elements in as pure a form as we can obtain them. We may now combine several elements into meaningful blends.

Let us use all elements but the first, or, in other words, let us combine the seasonal, the cyclical, and the irregular fluctuations. We shall obtain the Alpha production adjusted for trend. This would be the production if no growth factor were present.

Of more practical importance, however, is the combination of all elements except the second. Thus we show the production of the Alpha Corporation under the assumption that there is no seasonal factor present. Statisticians speak in this case of data adjusted for seasonal variations, or deseasonalized data.

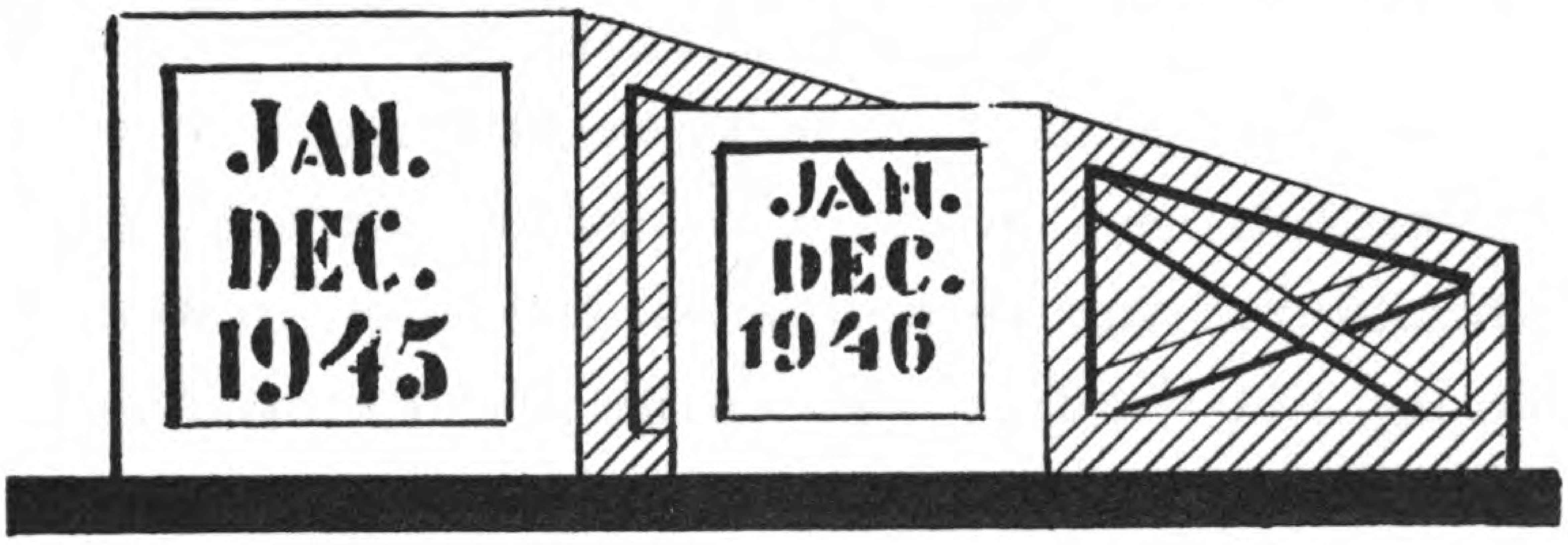
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

There are many instances where we need data of this type. If, for instance, we wish to compare the production data of the Alpha Corporation with those of the Beta Corporation, which has its seasonal peak in July, no meaningful comparison is possible unless we use deseasonalized data.

Another combination is important—the blend of the first and second elements. Trend and seasonal fluctuation together are called statistical normal. Any assumed monthly or quarterly production figure of the Alpha Corporation, based only upon the growth factor and the seasonal factor, represents the production that can be expected for a particular time period, or the normal production. Thus the Alpha production of next March takes into account the growth factor and the seasonal position of March.

If we deal with annual data, however, the expected or normal value is determined by the growth factor only. There is no seasonal factor in annual



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

data. If we lump together the twelve monthly production figures of one year, only the total annual production can be seen but no seasonal variation.

Thus, if we deal with annual data, only three of our four elements are present. Monthly, bimonthly, or quarterly data, however, consist of all the four elements that we described.

Our production data, observed in their variations over time, are a time series. Other series, such as sales, employment, prices, consumption, are frequently analyzed in their development over time.

The subjective element, which is generally present in statistical work is especially strong in time series analysis. The statistician decides about the methods of separating and measuring the factors that influence a time series; sometimes it is even up to him to classify a particular variation in the data. There may be doubt, for instance, whether it is due to cyclical or to irregular influence. The statistician will make decisions based upon his knowledge of the economic background of the particular time series and upon his experience in statistics. Different statisticians may arrive at different answers. This limitation of statistical analysis has to be viewed in the light of the fact that we are attempting here to measure economic concepts. Such concepts as trend

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

or cycles never can be measured with precision. We cannot, therefore, expect more than approximations in time series analysis.

Before we close our description of time series, let us give some attention to a problem which the statistician of the Gamma Corporation had to face.

Retail sales of the corporation in May and June of a certain year showed a considerable drop as compared with the preceding months. The sales manager was concerned about this and thought that the market might be saturated and Gamma sales would from then on show a tendency to decline; or, possibly, the Gamma Corporation would approach a phase of depression. Thus, he attributed the drop in sales to a change in the direction of the trend or to a change in the phase of the cycle.

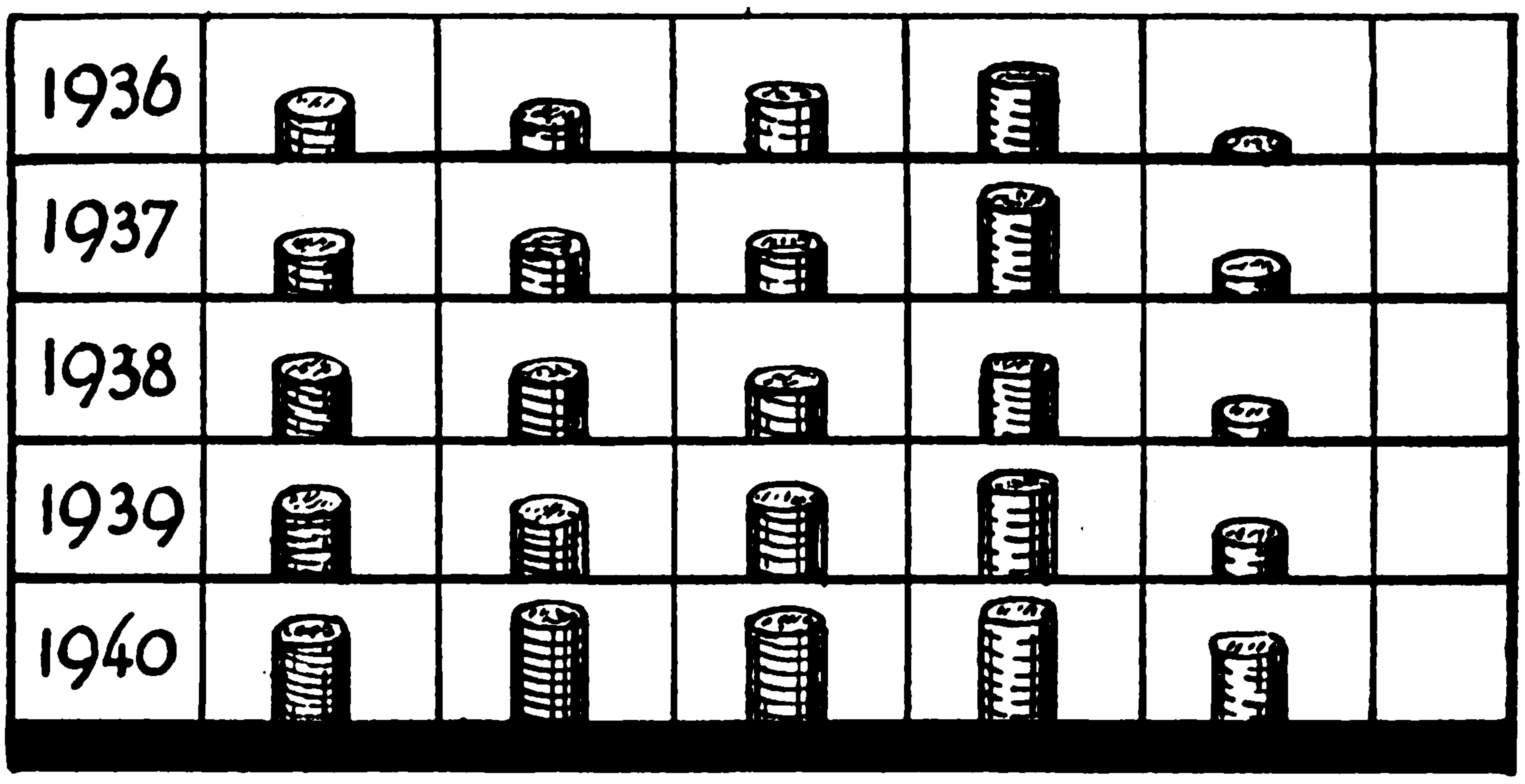
The statistician of the Gamma Corporation, however, pointed out that May and June sales of the corporation were always lower than sales figures of the preceding months. They were still lower than anticipated because there had been a short strike in one store and another store had to be closed for a few days because of alterations. Thus, seasonal variations and irregular factors may explain the low sales figures in May and June.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Index Numbers

The Alpha Corporation wishes to observe changes in the price level of their products. They manufacture five different articles, and the prices per unit show these changes during the period 1936-1940:

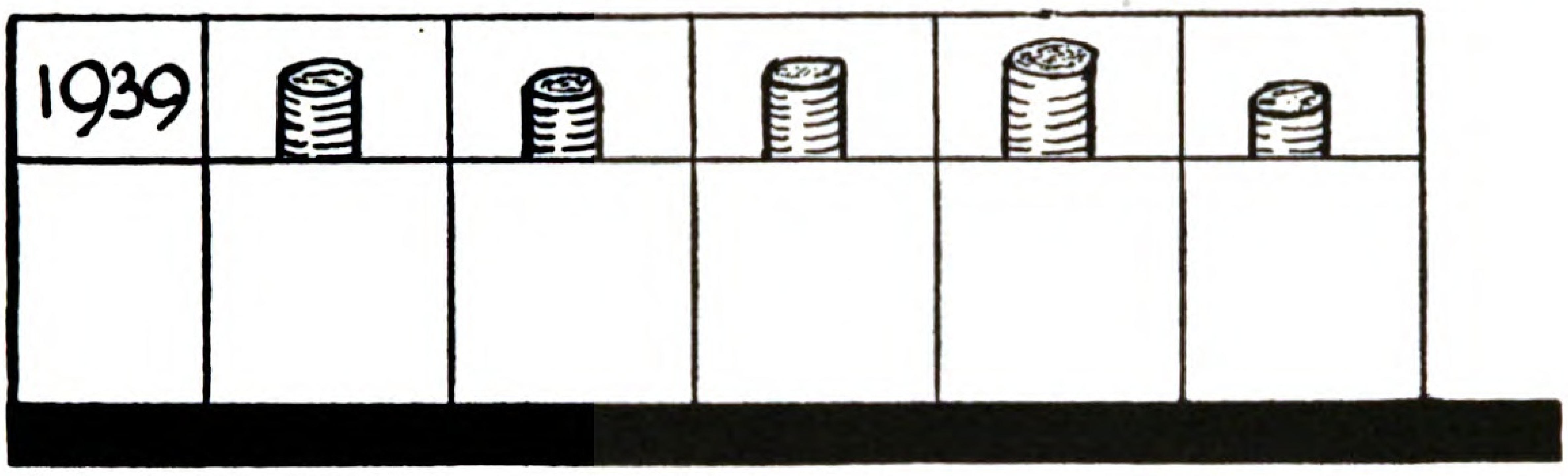


It is obvious that it would be extremely difficult to make a statement about the year-to-year changes of the price level. Again, statistics will help us to

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

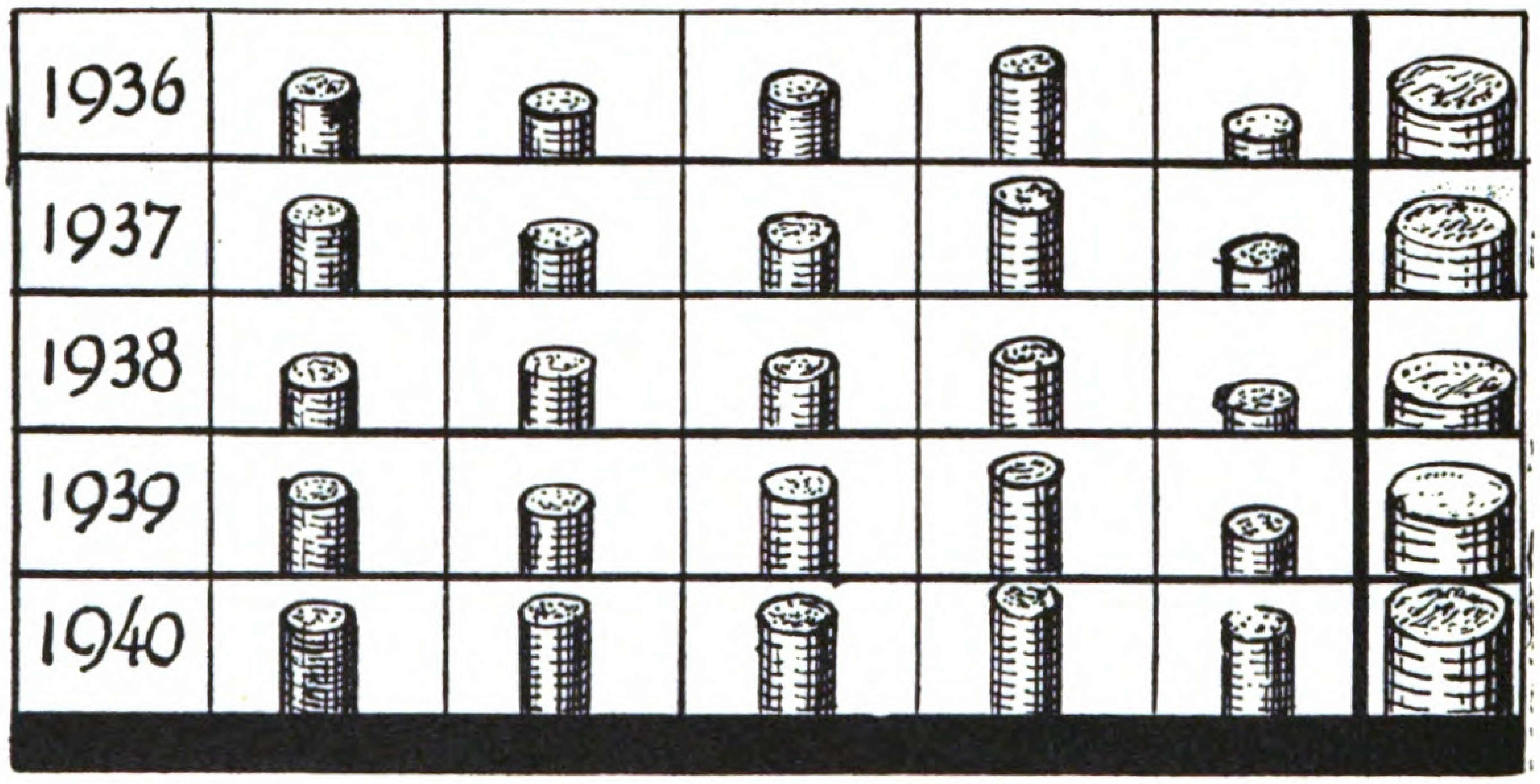
simplify and condense this information and thus facilitate comparisons.

We shall have to find a simple expression that combines the information pertaining to the five different items in any year.



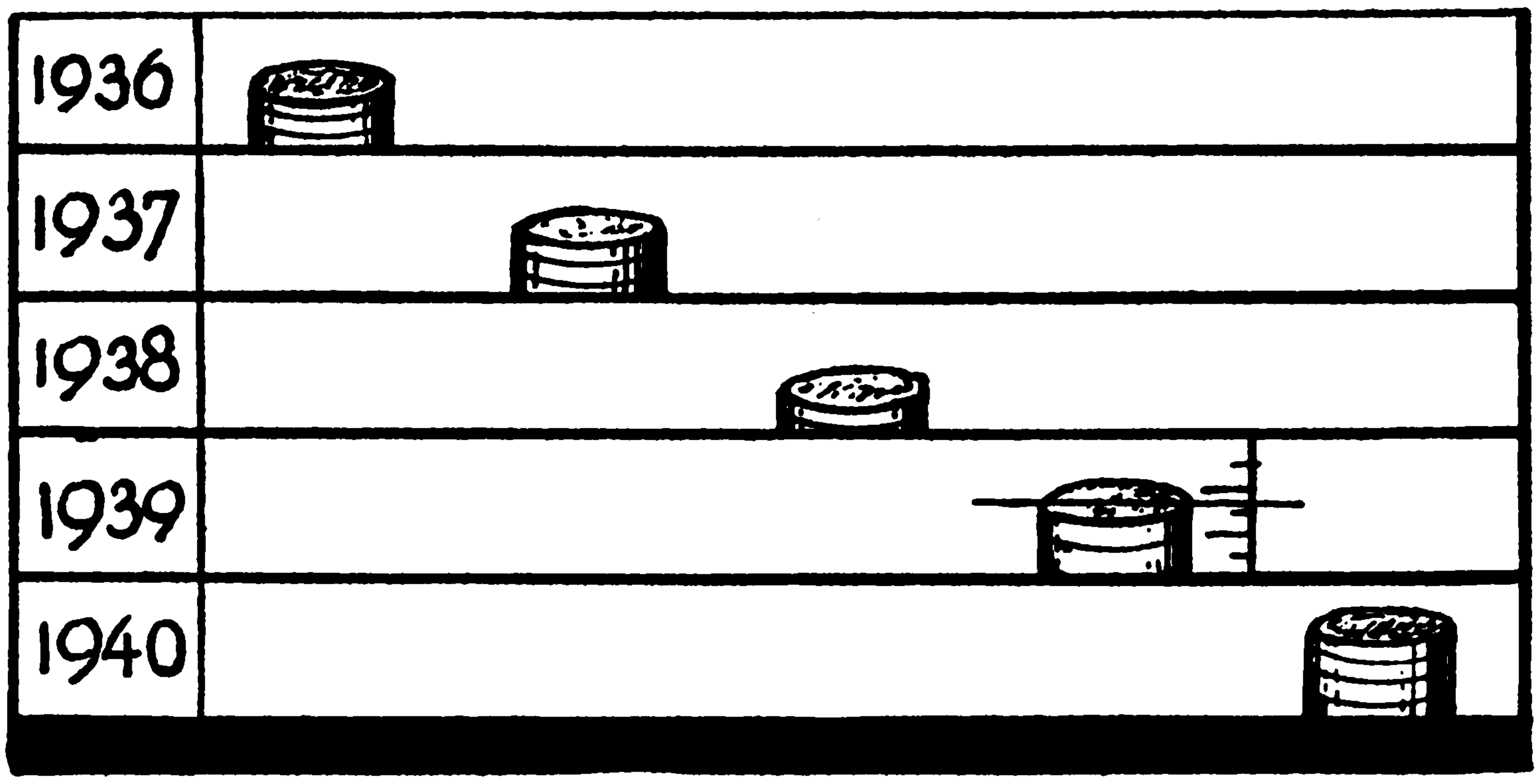
Thus we shall finally obtain a series of such expressions, one for every year, which are the resultants of the price changes taking place in the five articles.

We may combine the prices by totaling them for every year

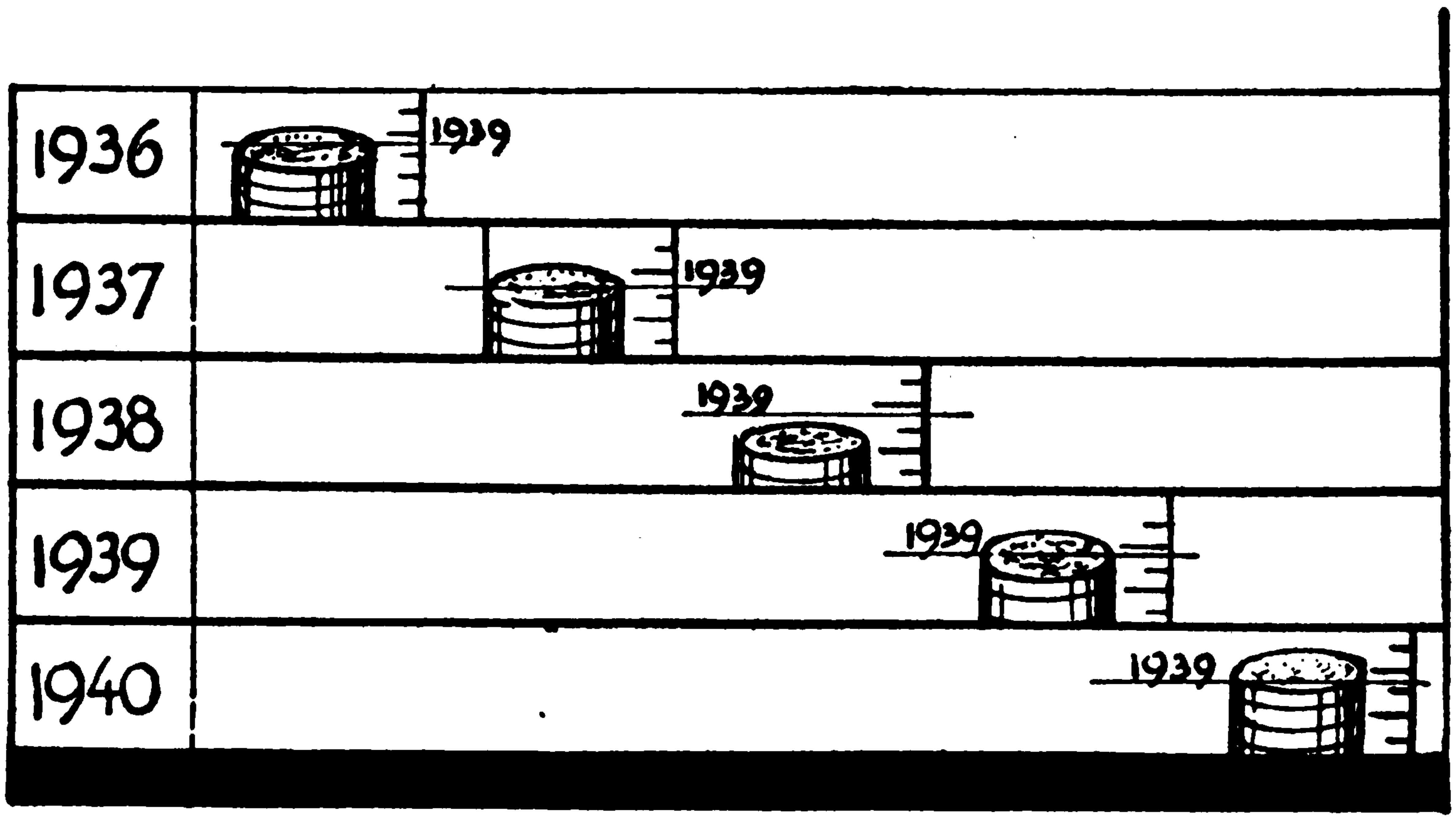


Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

and then we may, if we wish, make the price total of one year that we consider normal, the yardstick with which we measure all other totals. This yardstick, say 1939, is our base.



If we now apply our yardstick, we find that the 1938 total, for instance, is smaller than our yardstick, or, in other words, the price total in the base year is larger than the price total in 1938.



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The comparison of all data with the base will give us a simple series of measures that show us the changes in the price level of the Alpha products over a period of time. The price level in 1939 remains our base of comparison.

This is called an aggregative price index.

It so happens that the fourth of these commodities is priced per hundredweight and the others per pound. Naturally, the price of the fourth commodity dominates the price index and any change in the price of this commodity influences the index greatly. The reason for this predominant importance of the fourth commodity is the very superficial circumstance that it is quoted per hundredweight.

We shall avoid this unwanted emphasis by purposively applying an emphasis that brings out the

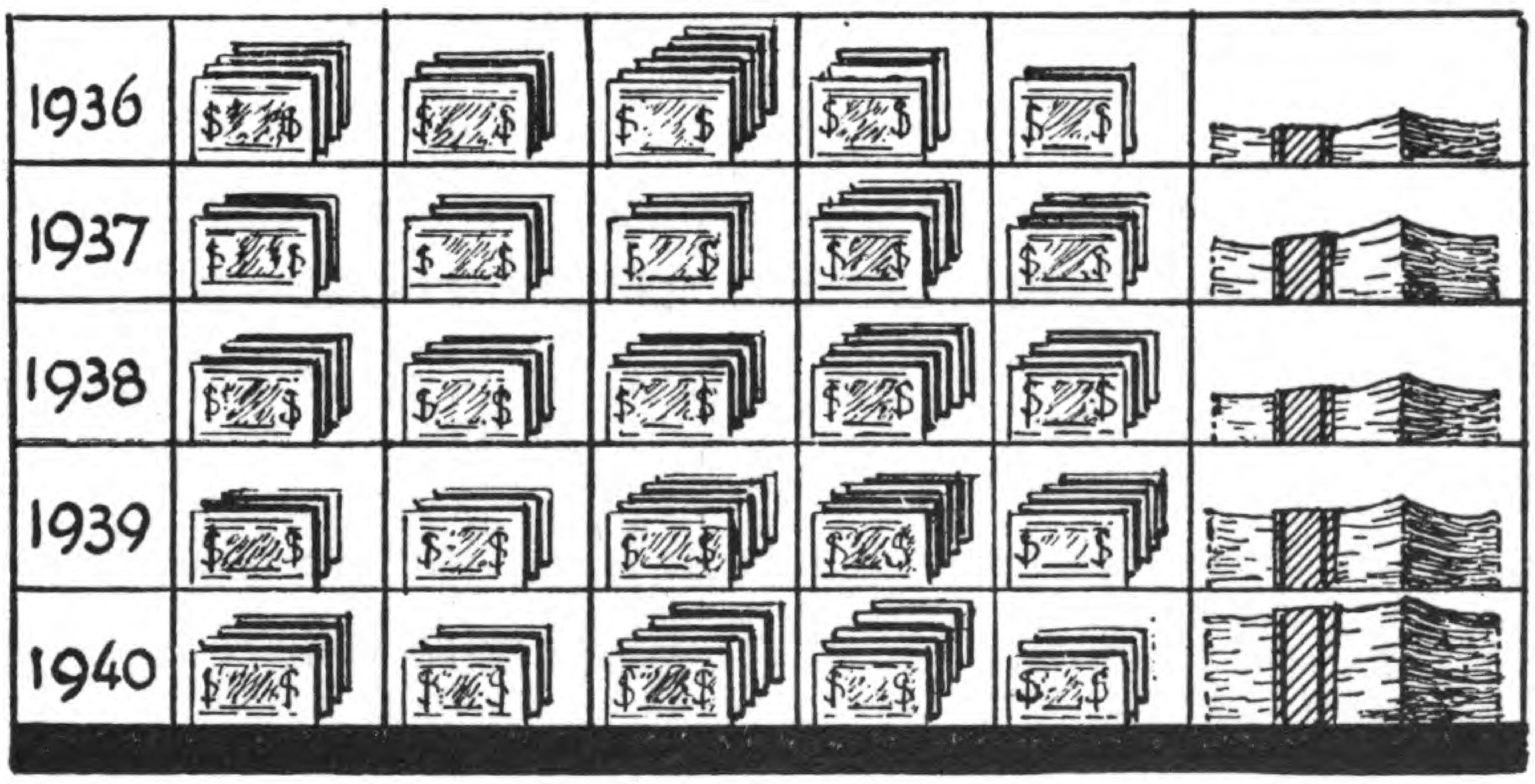
Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

relative importance of the five commodities for the Alpha Corporation.

Generally, the quantities produced, consumed, or marketed, taken at the base year or any other time period, will serve equally well the purpose of applying a sensible emphasis, or weight, as the statisticians call it.

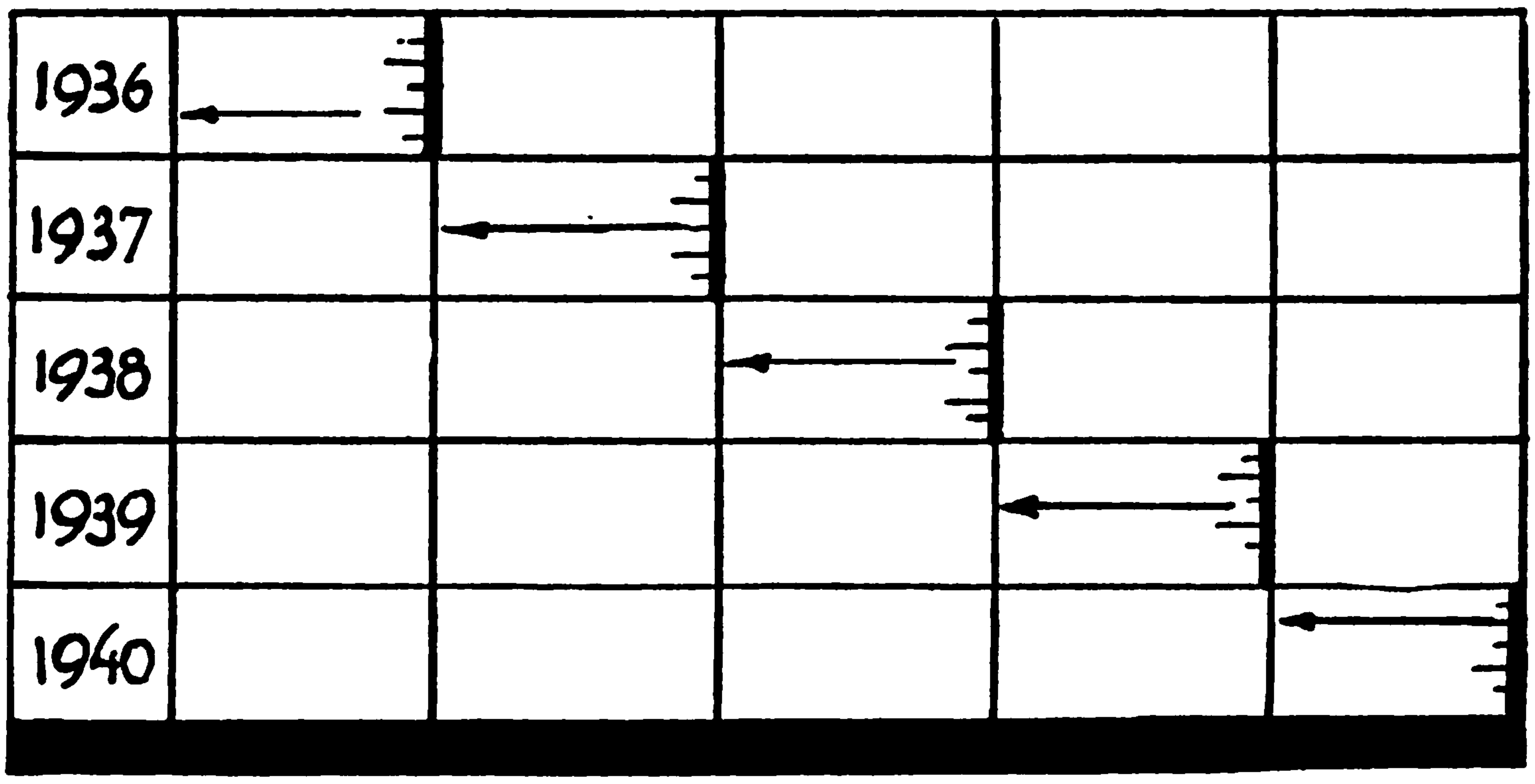
Let us take as a weight the quantity of each commodity sold during the base year. If we combine the quantity of the first commodity sold and the price at which this quantity was sold, we obtain the total value of the business done by the Alpha Corporation in this particular commodity. The same applies to the second, third commodity, and so on.

Now we must combine values.



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Again, one of these value totals will be used as a yardstick and we shall obtain a series of measures that will give us a better picture of the price changes since prices have been weighted according to their significance.



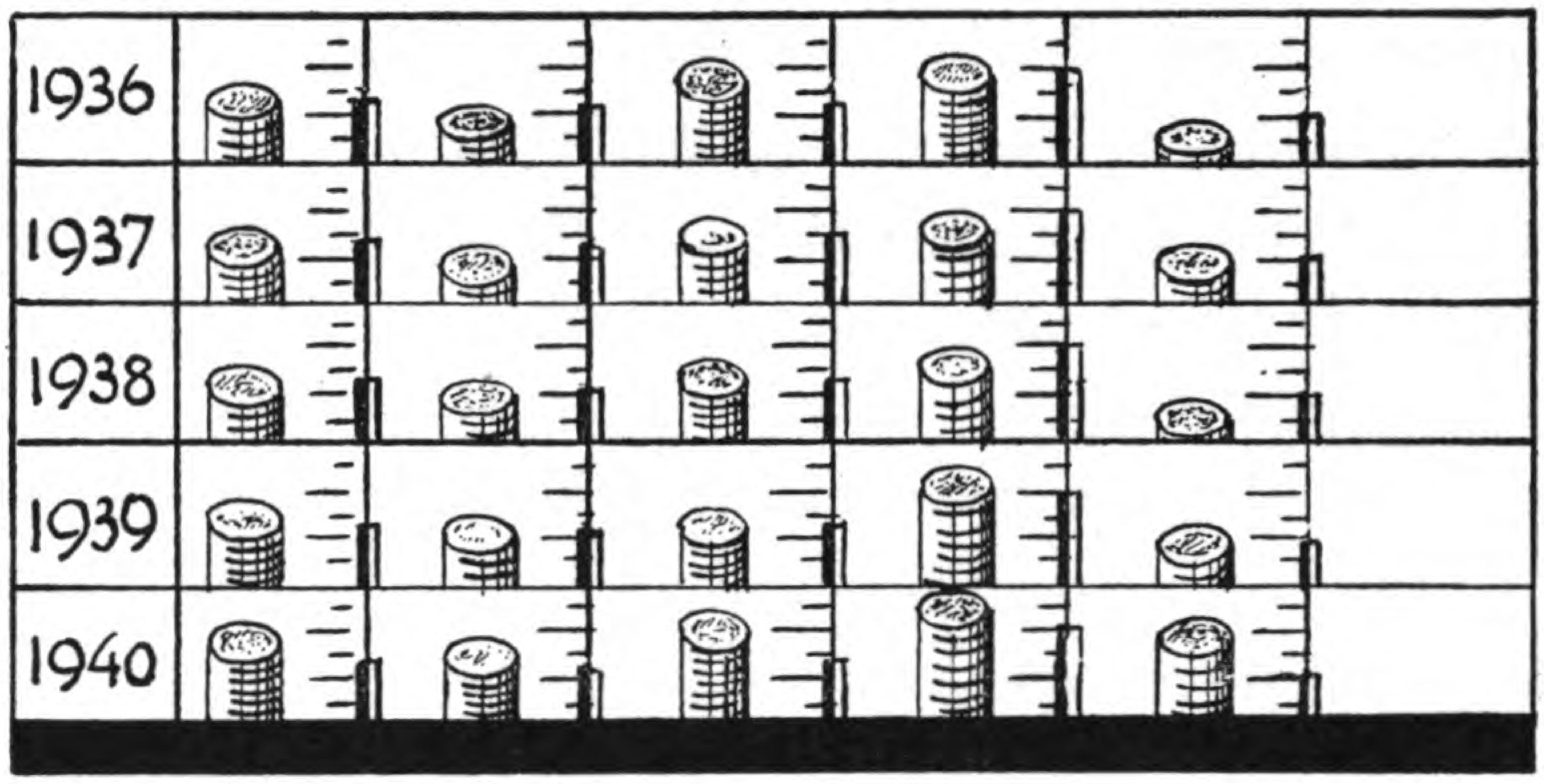
This is called a weighted aggregative index.

Some index numbers are of the average-of-relatives type. A different approach is used to obtain that simple series of index numbers.

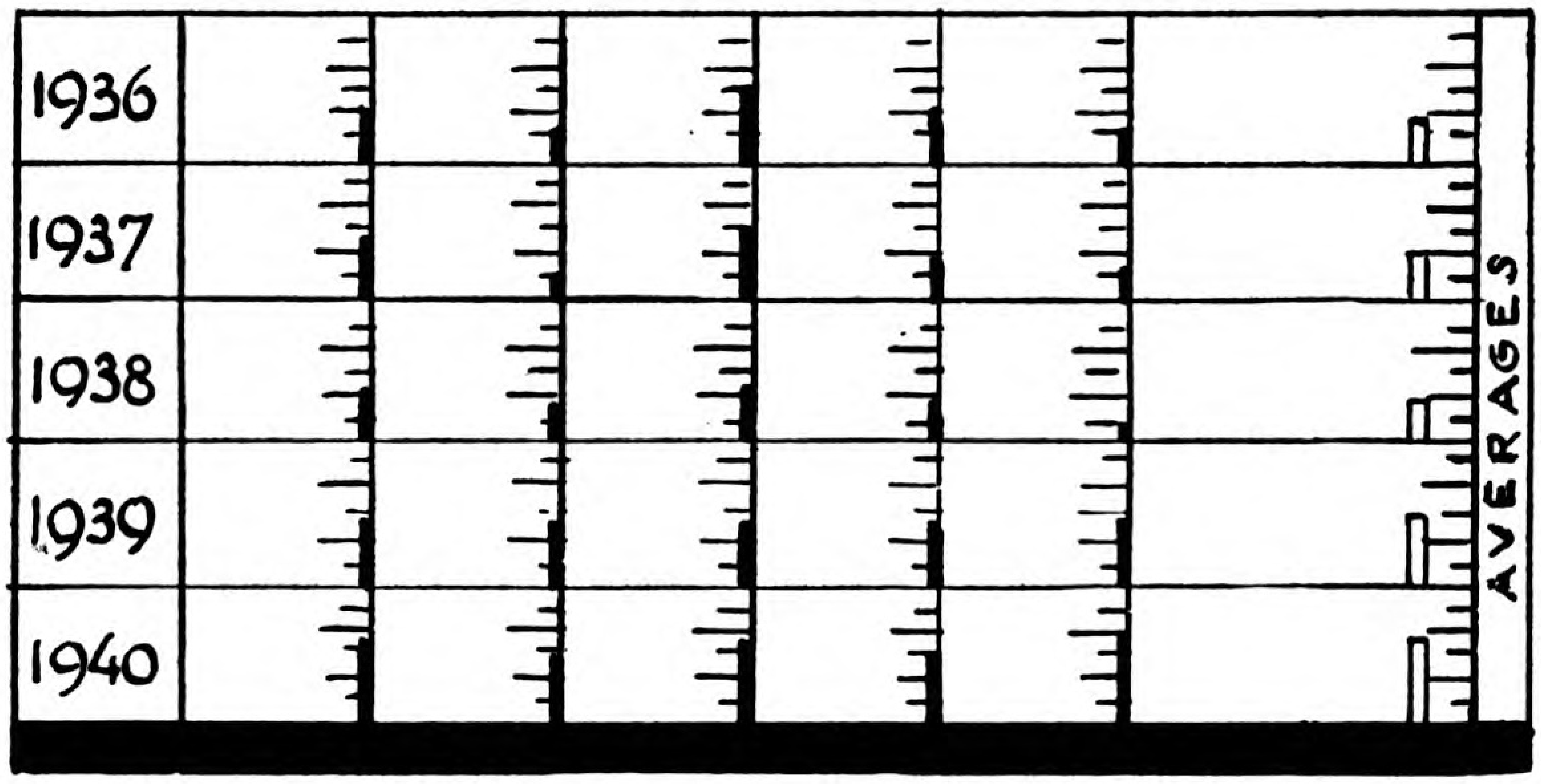
The price of every commodity in the base year is used as a yardstick to measure the prices of the same commodity in other years.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN



We obtain series of such measures and we represent every year by the average—any average of these measures.



Again, illogical weights are present unless we purposely apply weights.

We have compared changes in the price levels from year to year. We could also have compared

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

month-to-month changes, day-to-day changes—in short, changes between any time periods. The device of index numbers can also be used to compare price changes from place to place, where the prices in any given locality may be taken as the base.

There is a great variety of price indexes measuring not only changes in wholesale prices but changes in retail prices, cost of living, etc.

Everything that has been said about price indexes applies also to a second important group of indexes, which measures changes in quantities. They are called quantity indexes of production, consumption, etc.

We are comparing changes over a period of time in most of the index numbers, as we did in the time series analysis. In combining data, the concept of averages is important. In selecting the price or quantity data that are to go into an index we are very often selecting a sample.

Thus, certain limitations attached to the analysis of time series, averages, and to sampling are present in index numbers.

We shall, for instance, have to consider the effect of seasonal variations in dealing with monthly quantities produced. Furthermore, we shall remember that averages alone do not tell the full story. A statement concerning the variation about the aver-

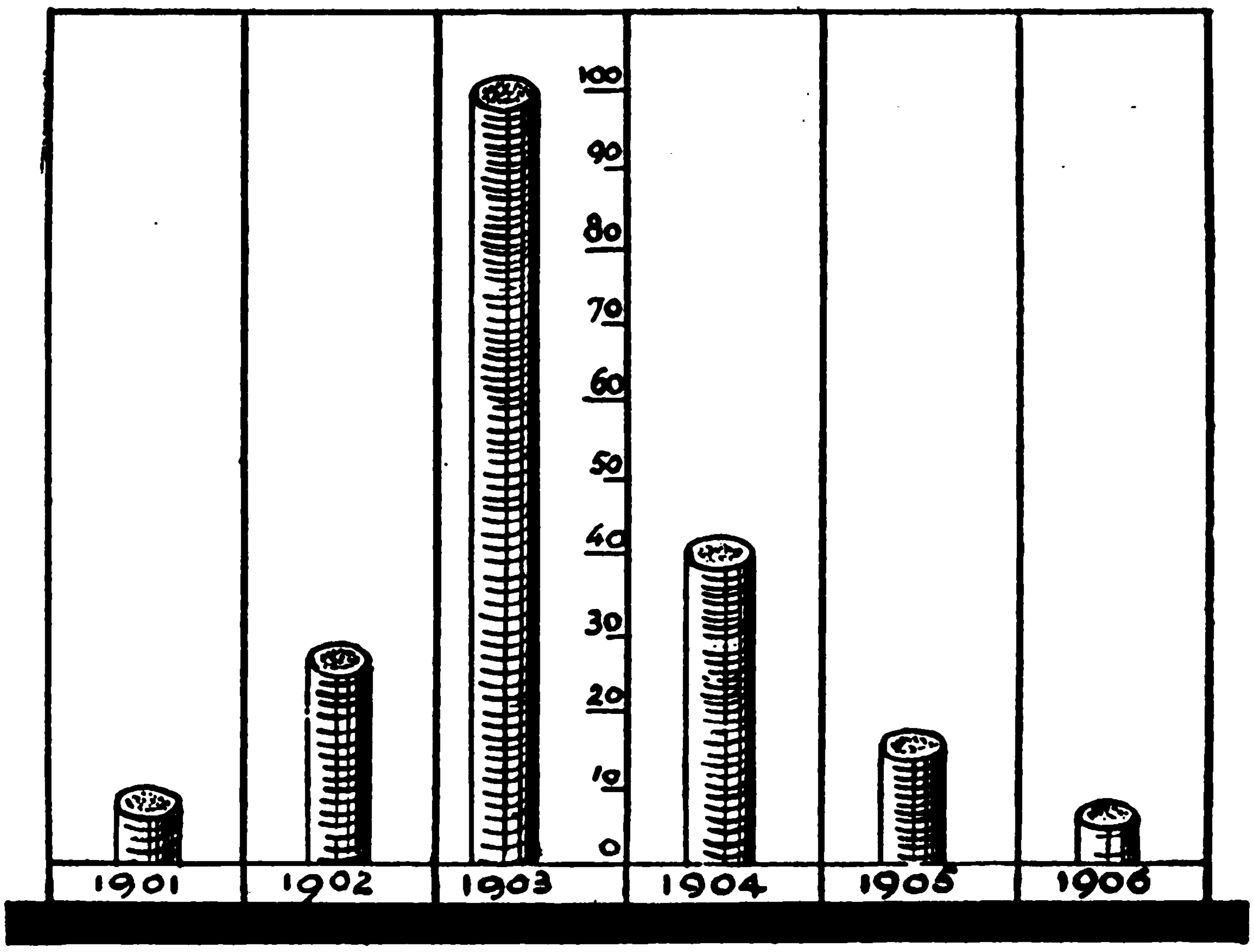
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

age is a necessary supplement. This means that an equal price level in two years does not give us any information about changes in the components.

Various aspects of sampling will be discussed later.

In using index numbers we are judging developments against a base. If the base is very high, all other index numbers will appear very small in comparison.



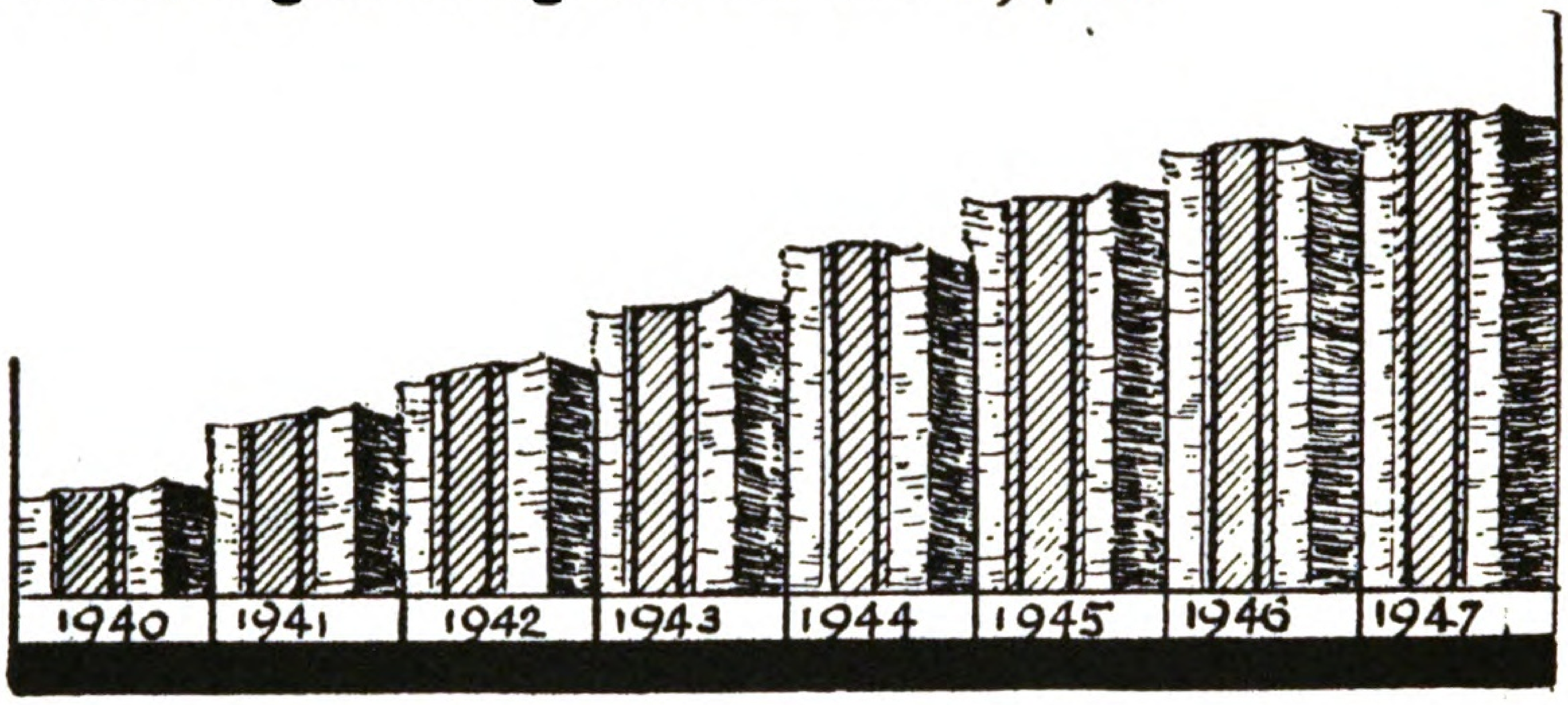
The opposite distortion will take place if a very small base is used for comparisons.

Index numbers are important instruments for

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

measuring changes from time to time, and, sometimes, from place to place. A special situation in which price index numbers are useful is the adjustment for changes in the Dollar value.

The Gamma Corporation shows, in Dollars, the following sales figures for the 1940's:



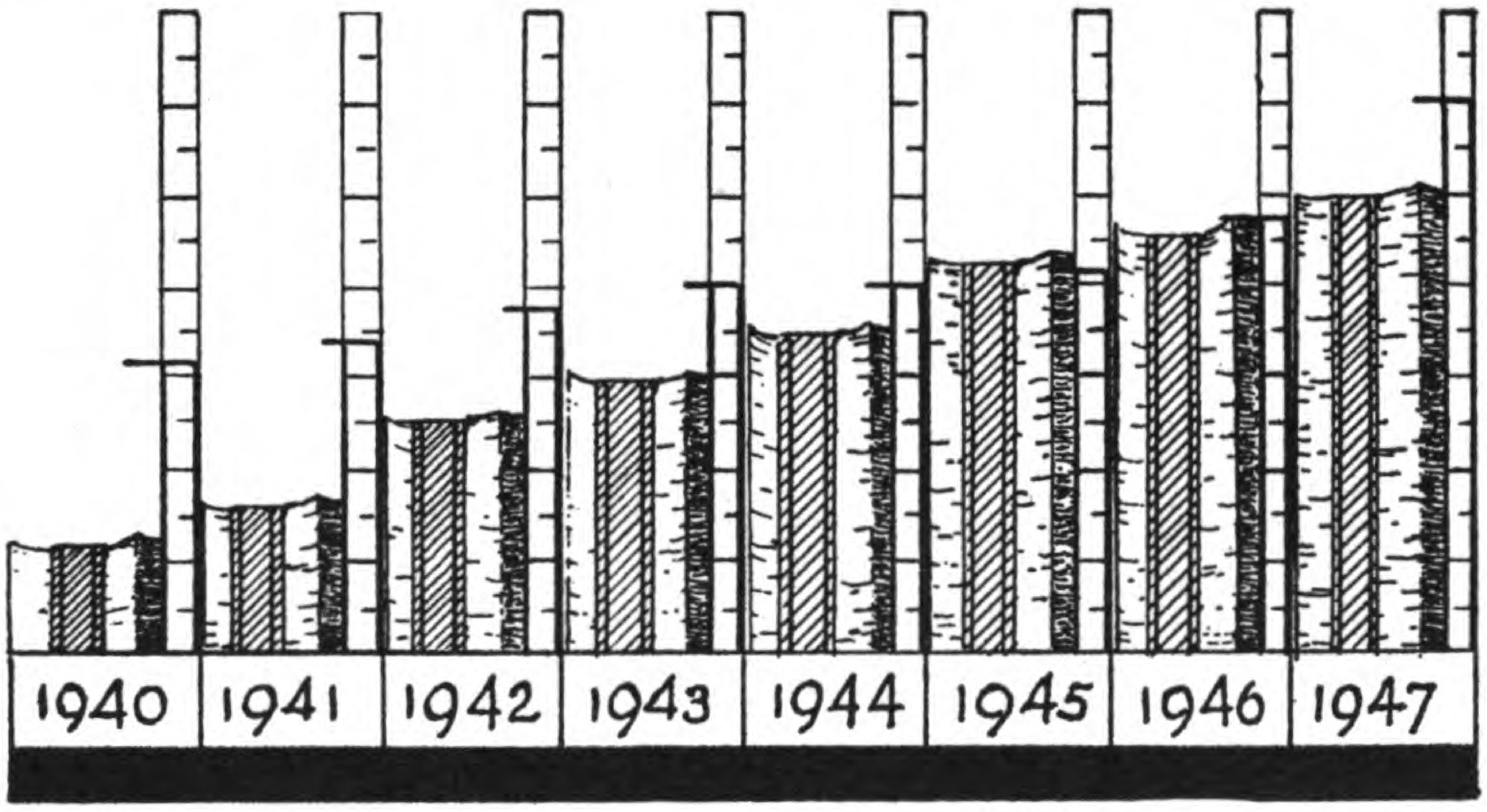
Does this mean that the volume of sales, the quantities sold, went up to the same extent?

We shall have to consider the prices at the different time periods. This will show us how far increased prices and how far increased quantities of merchandise sold were responsible for the increase in Dollar sales.

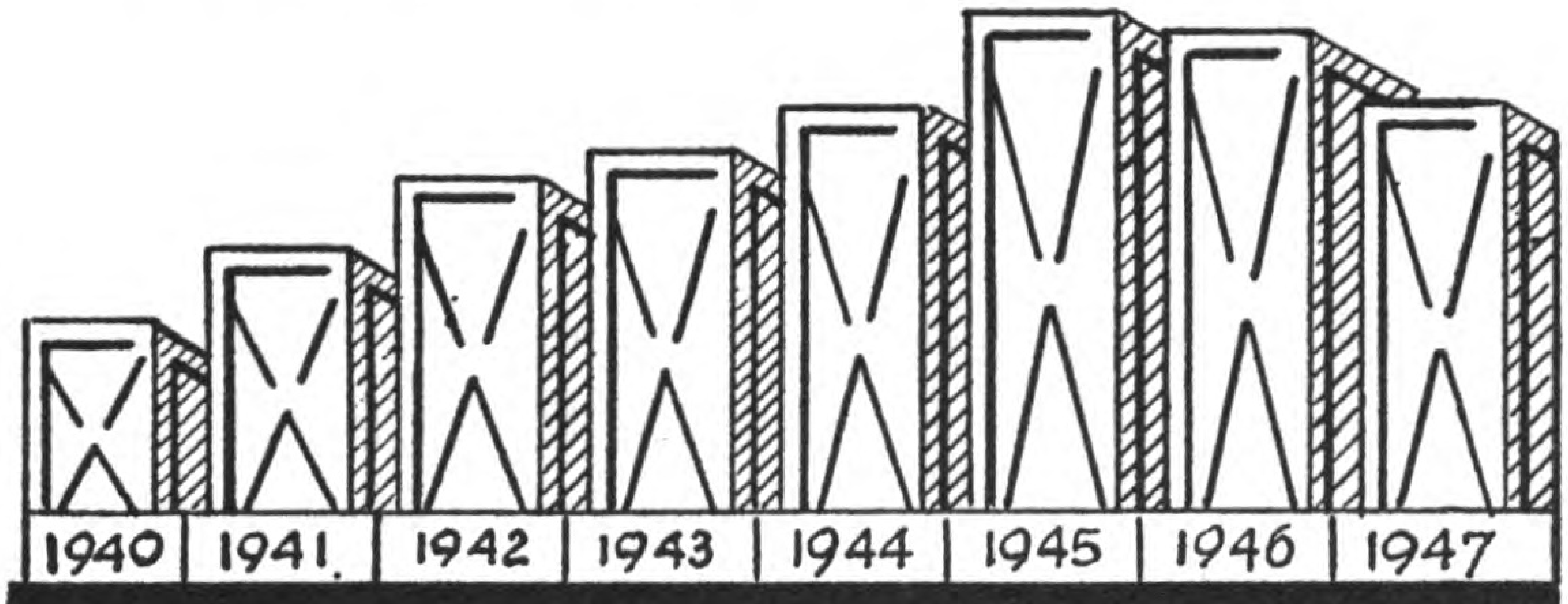
Changes in the price level are measured by index numbers. We shall select an index that meets the situation of the Gamma Corporation most closely. Since prices of this corporation are wholesale prices, we shall select a wholesale price index of the com-

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

modity traded by the Gamma Corporation, or, if not available, a commodity group that best fits the situation. The skill of the statistician will become apparent in the selection of a proper index. We shall compare this price index with the Dollar sales.



Statisticians call this process "deflation." It leads to the "real sales," that is sales based on the assumption that the Dollar value has not changed.



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Similarly, the personnel department of the Gamma Corporation may find the "real wages." These would be the hypothetical wages that result from a comparison of the actual wages and consumers' prices, or the so-called cost of living index. Real wages are wages based on the assumption that the cost of living has not changed.

Numerous indexes of prices, quantities, and of various other types are in use, mainly compiled by government agencies, research institutes, and magazines.

There is, for instance, the Wholesale Price Index of the United States Bureau of Labor Statistics. It is intended for the purpose of observing changes in the price level of the country by measuring changes in the wholesale cost of 889 important commodities. The index is constructed by the weighted aggregative method that has been described above.

Constructing an index of this type is a very difficult task and many limitations attached to such a measure have to be kept in mind by its users.

The Wholesale Price Index is sometimes used when it would be better to use one of its parts, an index referring to one commodity group only, like foods or textile products.

The base year of 1926 is too far in the past and

Digitized by Google | Original from UNIVERSITY OF MICHIGAN

the weights are based upon an economic pattern that is out of date.

Tremendous problems confronted the makers of this index during World War II. No fully satisfactory solution could be found, so that the utility of this index is reduced as far as it refers to the war years.

The same limitation applies also to another index of the Bureau of Labor Statistics, the so-called Cost of Living Index, or "Consumers' price index for moderate income families in large cities." It measures price changes in goods and services which are important to moderate-income families. This index, too, uses the weighted aggregative method.

An important index showing changes in the industrial production of this country is being constructed by the Board of Governors of the Federal Reserve System. This quantity index uses the weighted average of relatives method.

A great number of other index numbers have also become significant guide lines for economic thinking.

The following will illustrate how these indexes are used in practice:

Every month the statistician of the Beta Corporation compares the fluctuations of the Beta prices

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

with the Wholesale Price Index of the Bureau of Labor Statistics. He uses for the comparison the subgroup of this index which contains articles similar to those manufactured by Beta.

The Beta Corporation also compiles an index of production, showing the changes in the quantities produced by the corporation, adjusted for seasonal variations. Each index number is a percentage of production in the period 1935-1939. The statistician of Beta compares his own index with the most suitable subgroup of the Federal Reserve Index of Industrial Production, also adjusted for seasonal variation.

The so-called "Cost of Living Index" (Consumers' Price Index) of the Bureau of Labor Statistics is used to assist the Beta Corporation in adjusting wages of their workers in order to keep "real wages" constant.

Digitized by Google Original from UNIVERSITY OF MICHIGAN

Correlation

THE sales manager of the Alpha Corporation is fully aware of the fact that there is a relationship between the advertising done by its retail stores in different cities and the sales of these stores.

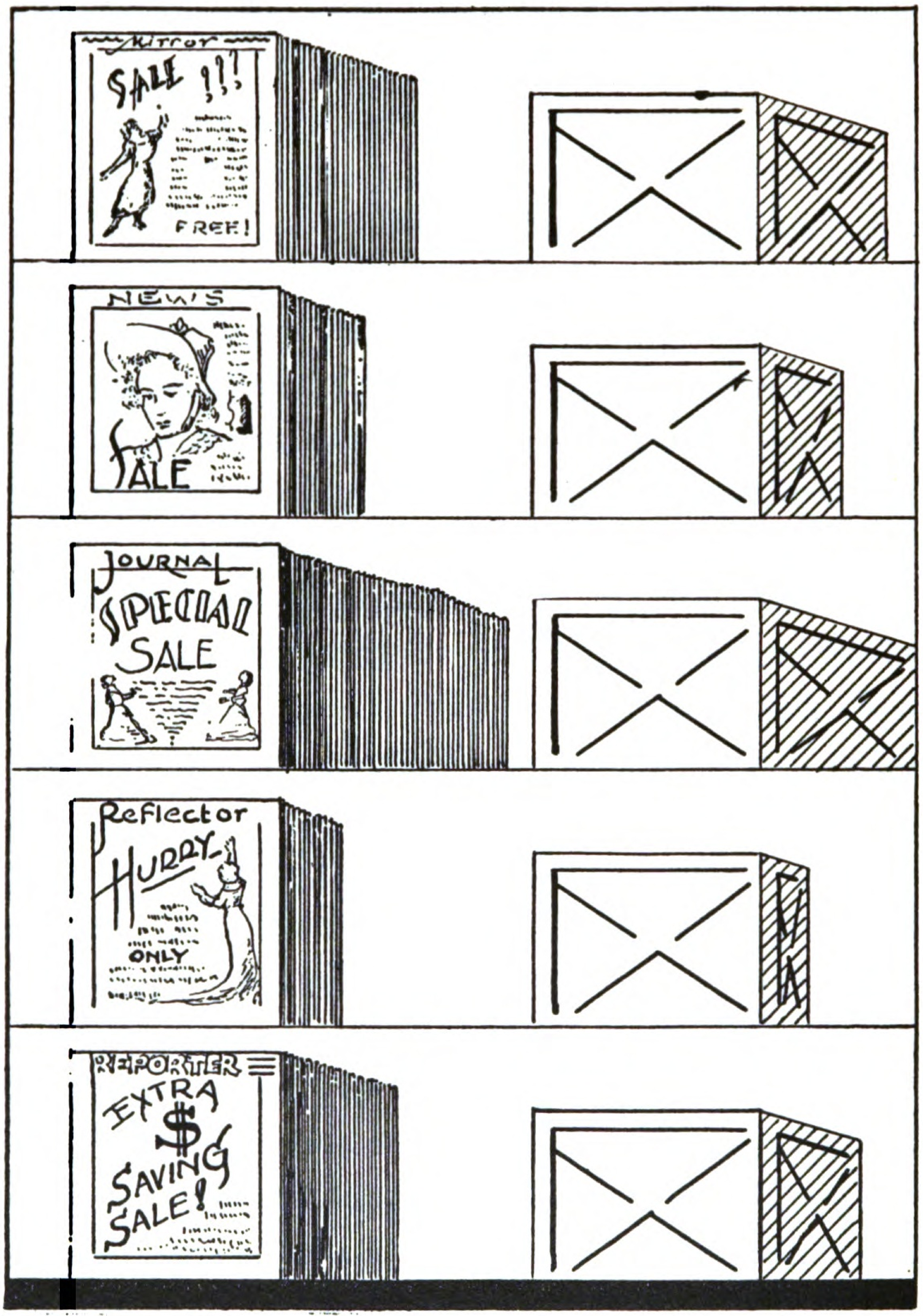
There is clearly an association between sales and advertisements. Greater advertising activity goes with larger sales, little advertising means smaller sales. Such a relationship is called positive correlation.

The purchasing department of the Alpha Corporation can supply a different illustration. Whenever a certain raw material, used by the corporation, is plentiful, prices are low.

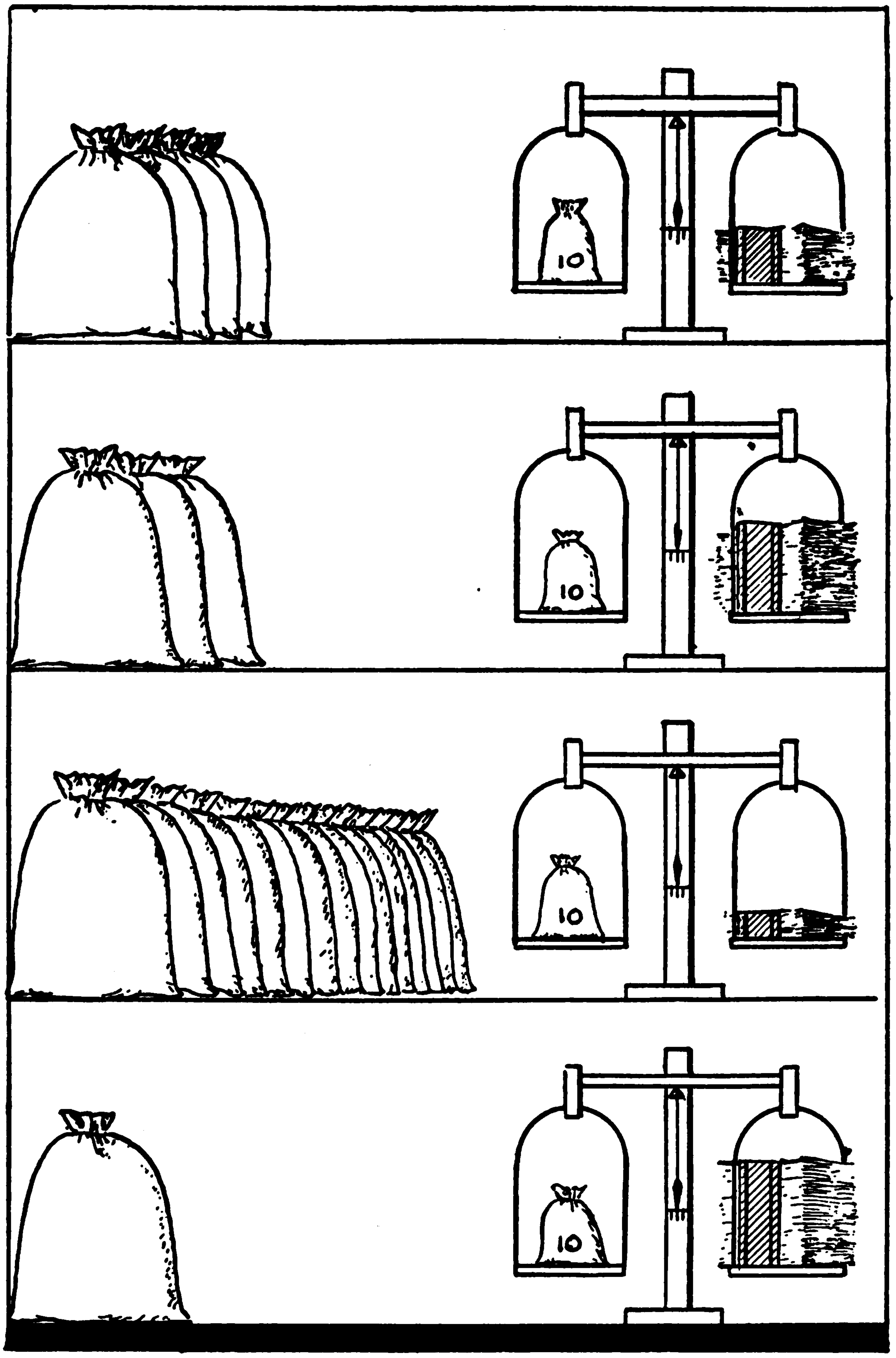
A large crop means a low price and a small crop a high price. This is a negative correlation.

If we were to analyze the relationship between advertising of the Alpha Corporation and the age of its employees, we would find no association. Neither

Digitized by Google Original from UNIVERSITY OF MICHIGAN



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

would we find high advertising figures to go with high age nor high advertising figures with low age. There is no correlation.

Statistics can establish the degree to which advertising and sales vary together or crops and prices vary in opposite directions. This is how far statistics alone can go. There is, however, the possibility of a more far reaching relationship. It may be found that sales, to some extent, depend upon advertising; this would be a functional relationship. Certain increases in sales may be caused by increased advertising—a causal relationship. To establish functional or causal relationships, tools other than that of a statistical nature must be used.

Thus, the mere statistical measure of correlation has to be interpreted in the light of logic and the background of the data on hand. We may for instance find a high degree of correlation between Alpha sales and the prices of the Gamma Corporation. Every time Gamma prices go up, Alpha sales go up. There is no functional or causal relationship between these two sets of data but both may depend upon, or may be caused by a third factor, the general business situation. Furthermore, the existence of a relationship may be entirely due to chance.

Our knowledge of correlation may be used for

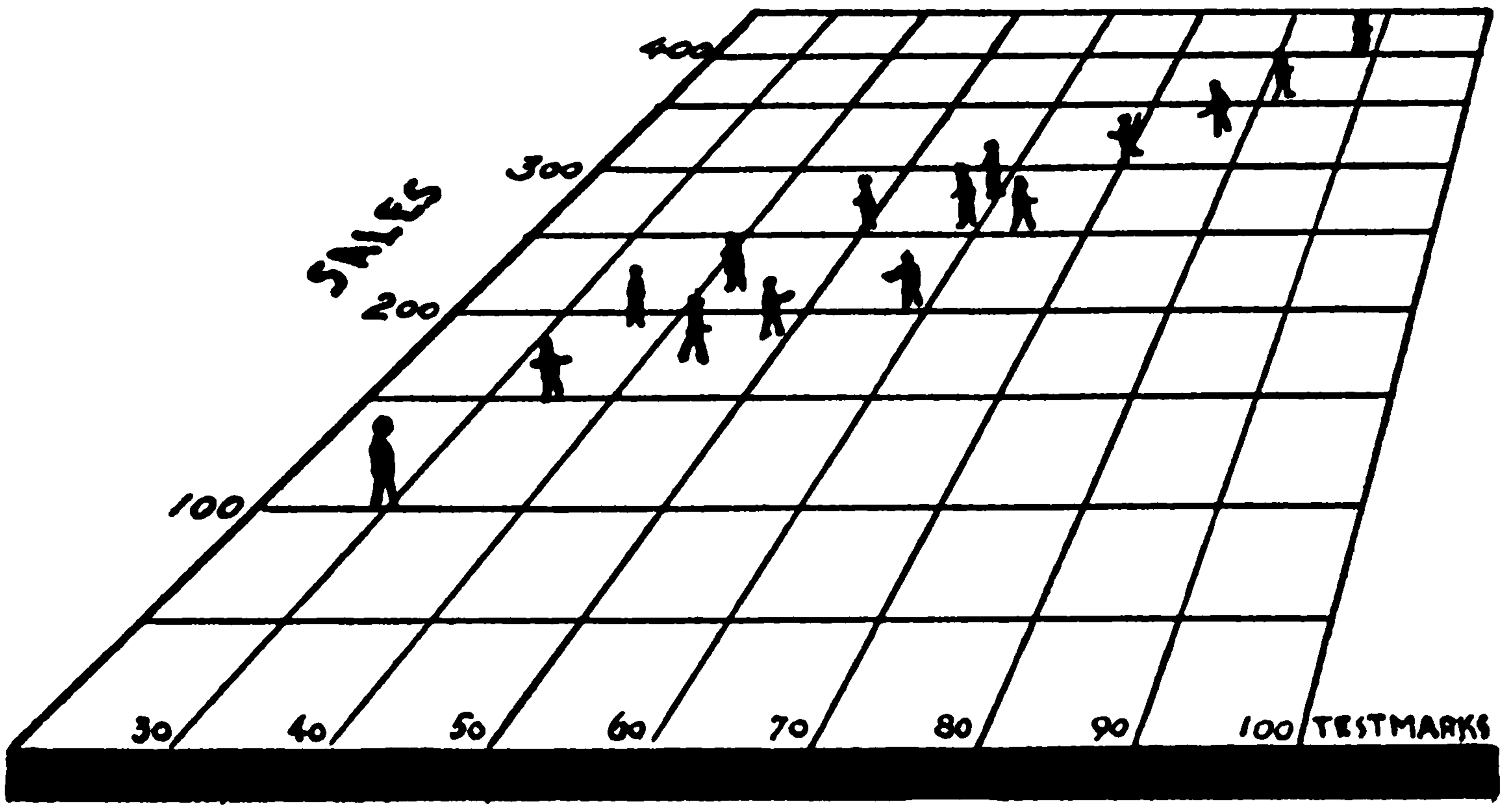
Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

predicting, which is one of its most important aspects.

Let us investigate the relationship between grades in an intelligence test which every Alpha salesman has to take, and his sales performance during the last month.

In the yard behind the office building we may designate certain spaces for the salesmen according to their sales and their test marks. We shall ask them to watch both sales and test signs when they look for their proper places.



We notice that the salesmen are standing along a path that runs roughly from the left corner in front to the right corner in the rear. This is a consequence of the fact that high test marks go with high sales

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

figures and low marks with low sales. In other words, there is a positive correlation between these test marks and the sales performance.

This arrangement of the salesmen along such a path gives us the basis for a rule of thumb: If a salesman has a certain test mark, he may be expected to show a certain sales figure.

This rule of thumb is very useful. For instance, when we get a new salesman, we may, from his test mark, approximately predict his sales performance. The test mark may even be higher or lower than the marks that occurred previously.

The narrower the path, the greater the association between test marks and sales; the greater this association, the more reliable will be predictions based upon this association.

The path can be described by a line, the so-called regression line. This term comes from the first use of this concept in a biological study; however, it has no connection with most of the situations in which it is now used.

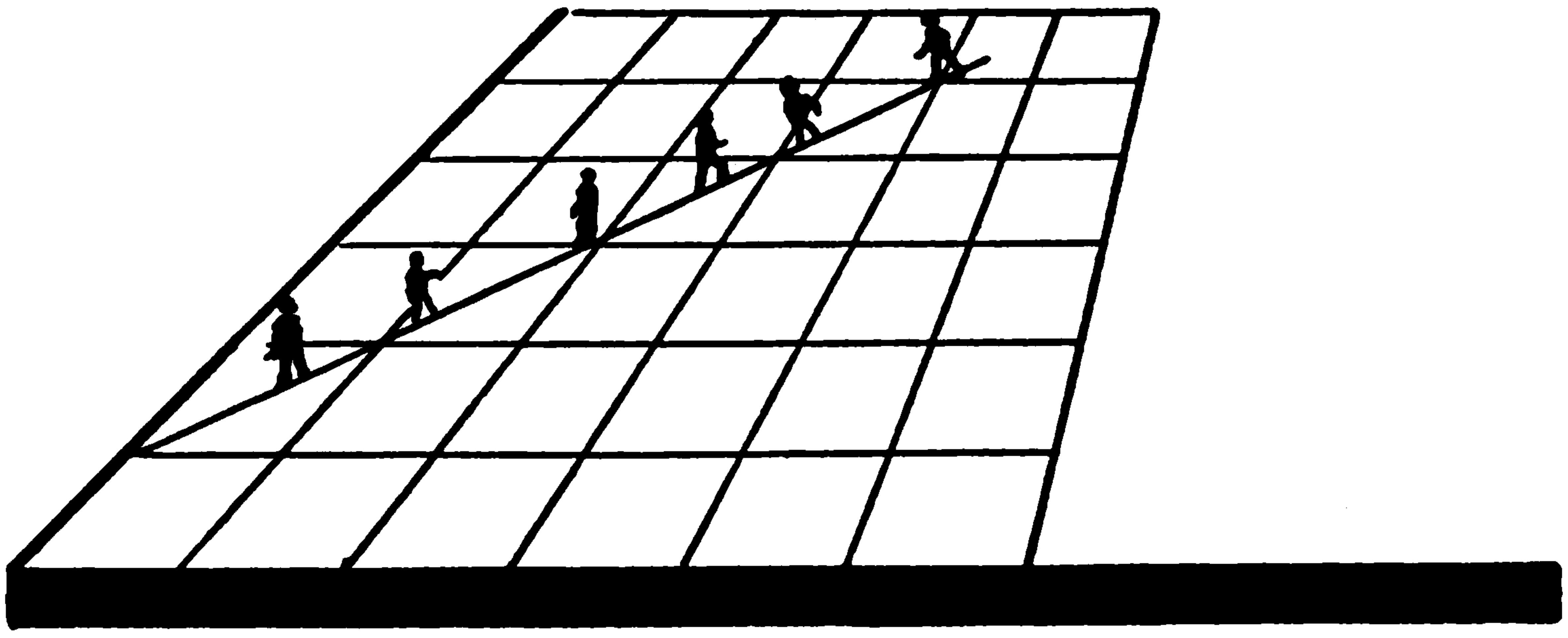
The reason for the slight scatter about our regression line is that the intelligence of the salesman, which is measured by the test, is not the only factor that determines his sales. Good appearance, pleasant manners, and other features may be of importance.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Were intelligence the only determinant of sales, we would find all the Alpha salesmen standing right along the regression line, not a path.

This situation of a complete relationship, which hardly ever occurs in practice, is termed perfect correlation. Of course, positive as well as negative correlation can be perfect.



If we had found a perfect correlation, we could predict sales from test marks with absolute safety. Sales could be fully explained by test marks only.

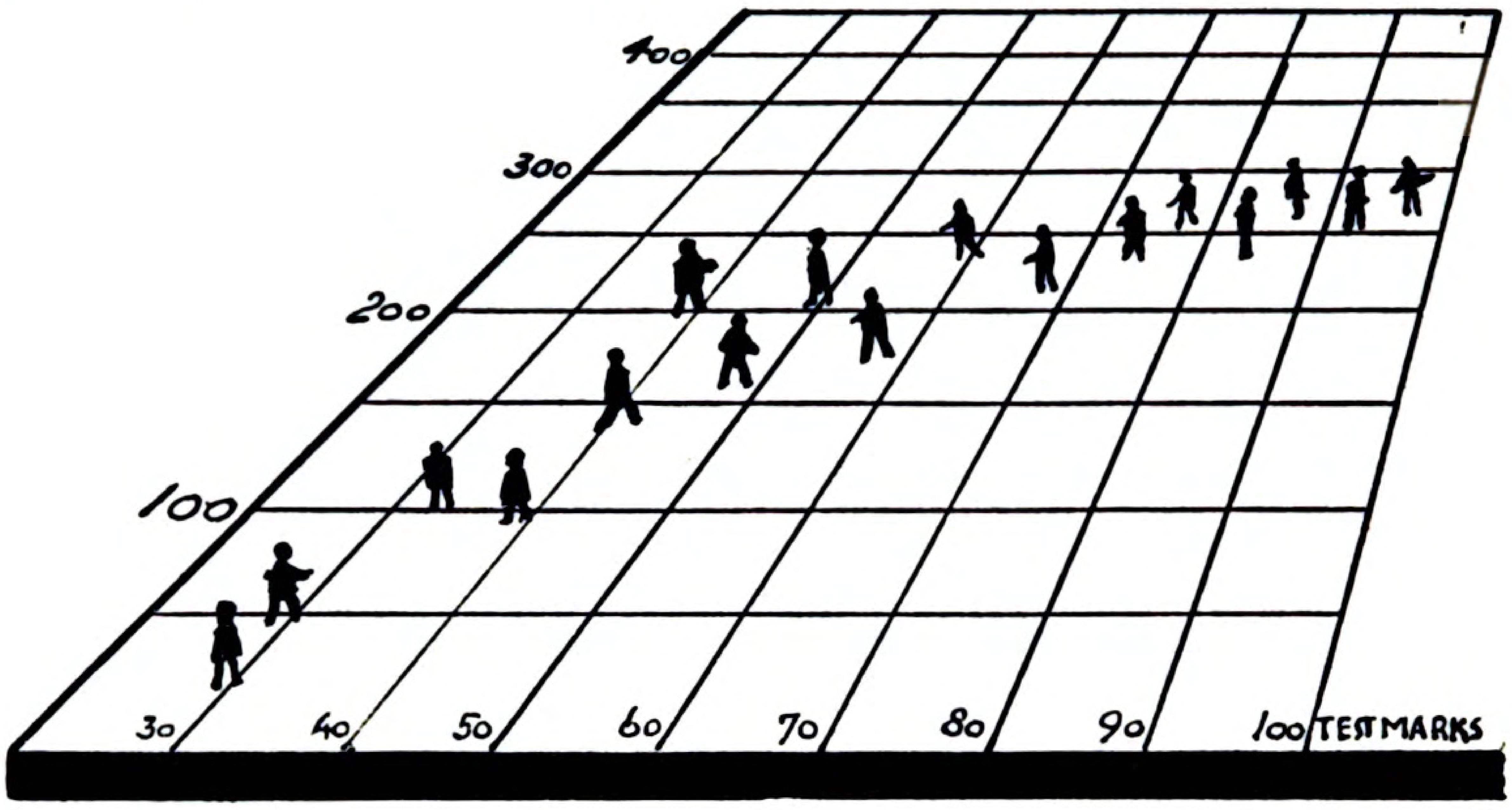
We saw before that the salesmen of the Alpha Corporation were standing on a straight path. This type of correlation is the linear correlation.

Salesmen of the Beta Corporation meet a different situation. They find it very difficult to sell more than a certain maximum amount, regardless of their abilities. If we would ask them to take their places according to test marks and sales, as we did with the

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Alpha salesmen, we would now see the following picture:



This clearly is not a straight but a curved path. It indicates a relationship that is designated as curvilinear correlation.

Linear as well as curvilinear correlation are types of simple correlation. The criterion of simple correlation is that we are analyzing the relationship between only two sets of data, as for instance test marks and sales.

A great number of other relationships are capable of analysis. To mention just a few illustrations, we may be interested in the relationship between national income and retail sales, or output and operating expenses, or earnings per share and price per share.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

The statistician of the Gamma Corporation, for instance, is interested in the relation of inventory to sales. Based on an average relationship he knows that a given sales figure may be expected to go with a certain value of the inventory. A larger or smaller value of the inventory would not conform with this norm.

Since the sales manager has an estimate of next year's sales, the expected value of next year's inventory may be predicted, based on the average relationship of inventory and sales.

Finally, the degree of relationship between inventory and sales can be measured by the so-called coefficient of correlation. When the value of this measure is high, the statistician of the Gamma Corporation may place great confidence in the statements and predictions that he based upon the relation of inventory to sales.

A limitation has to be mentioned that may be important in interpreting correlation.

We analyzed the relationship between advertising and store sales in different cities. Let us assume that these cities were all medium-sized provincial towns and we had the information referring to the large New York store. Advertising as well as sales are much larger than in the other places. The measure

Digitized by Google

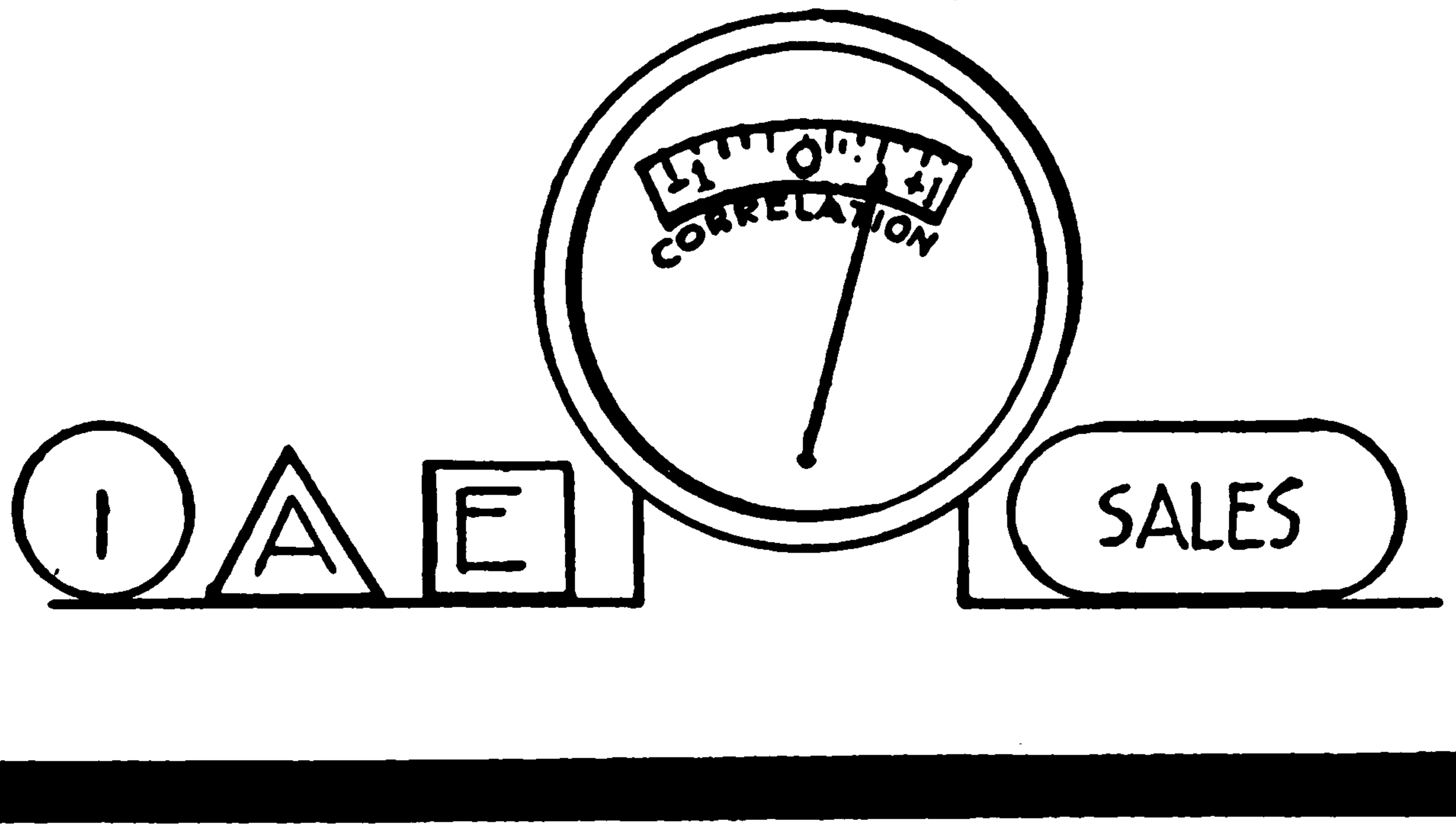
Original from
UNIVERSITY OF MICHIGAN

of correlation between advertising and sales would be influenced by the extreme New York data to such an extent, that it might not depict the entire relationship properly.

Other Types of Correlation

We mentioned before that several factors influence the sales of an Alpha salesman—his intelligence, appearance, etc. Let us select three of them: intelligence, as measured by the test, his age, and his experience in years of work as a salesman.

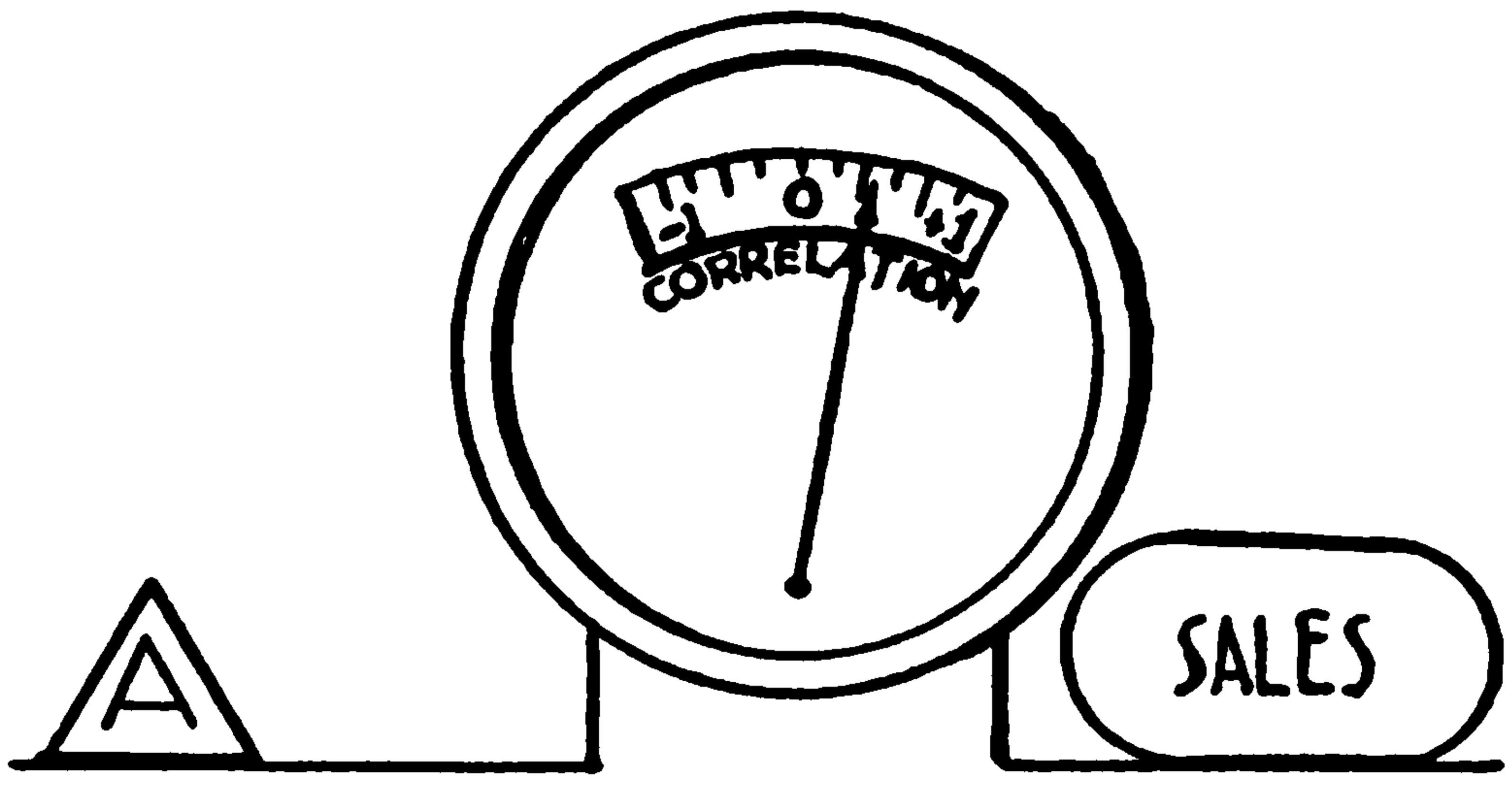
We may study the combined effect of these three factors upon the sales and by doing so we are using multiple correlation.



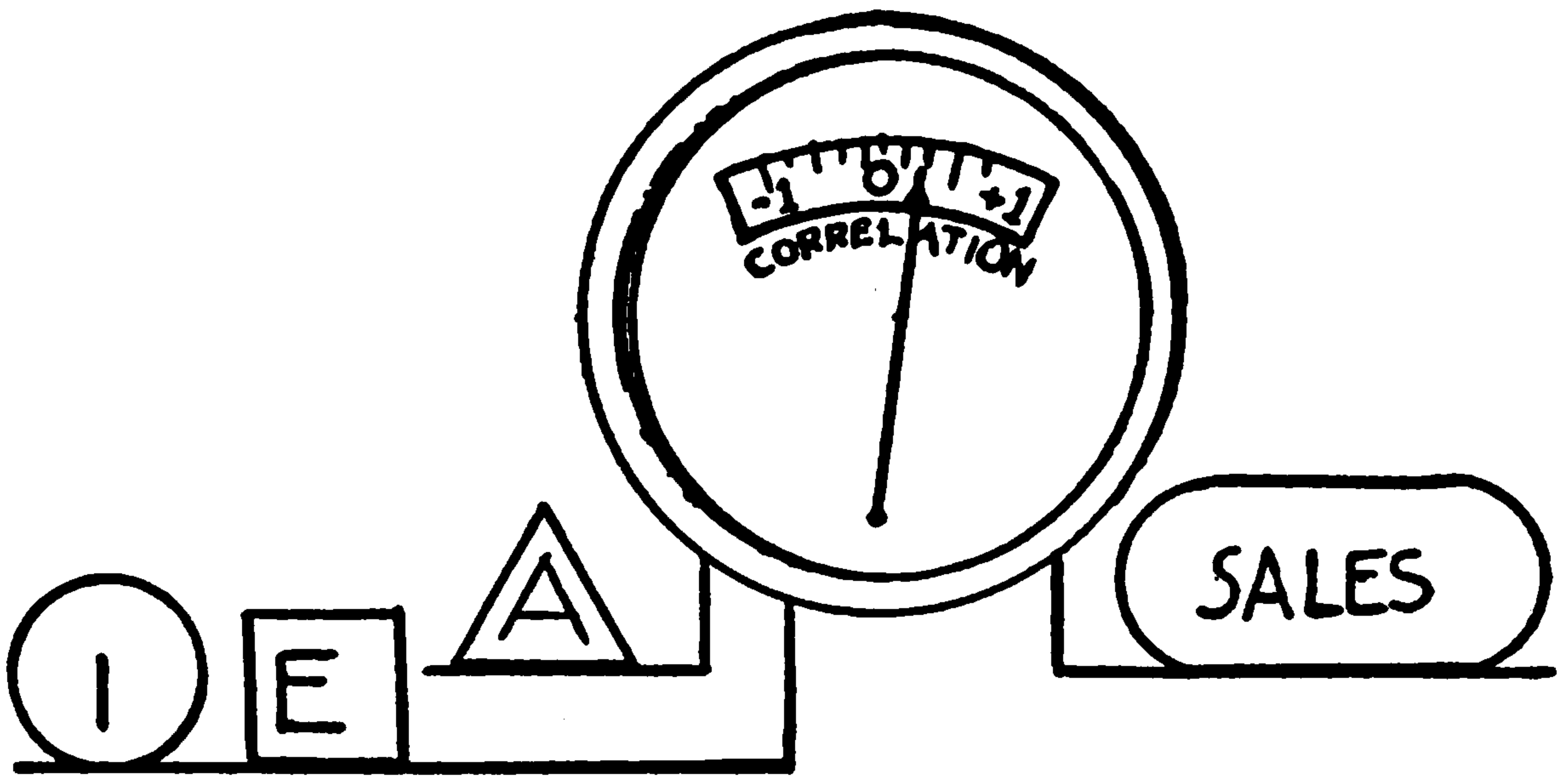
However, we may just be interested in the relationship of, say, age and sales. If we consider this relationship, the problem is one of simple correla-

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

tion, which was discussed before. We are here disregarding the influence of intelligence and experience.



Should we, however, study the association between sales and age but at the same time take into consideration the relationship existing between sales on one hand and intelligence and experience on the other hand, the problem becomes one of partial correlation.



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Sampling

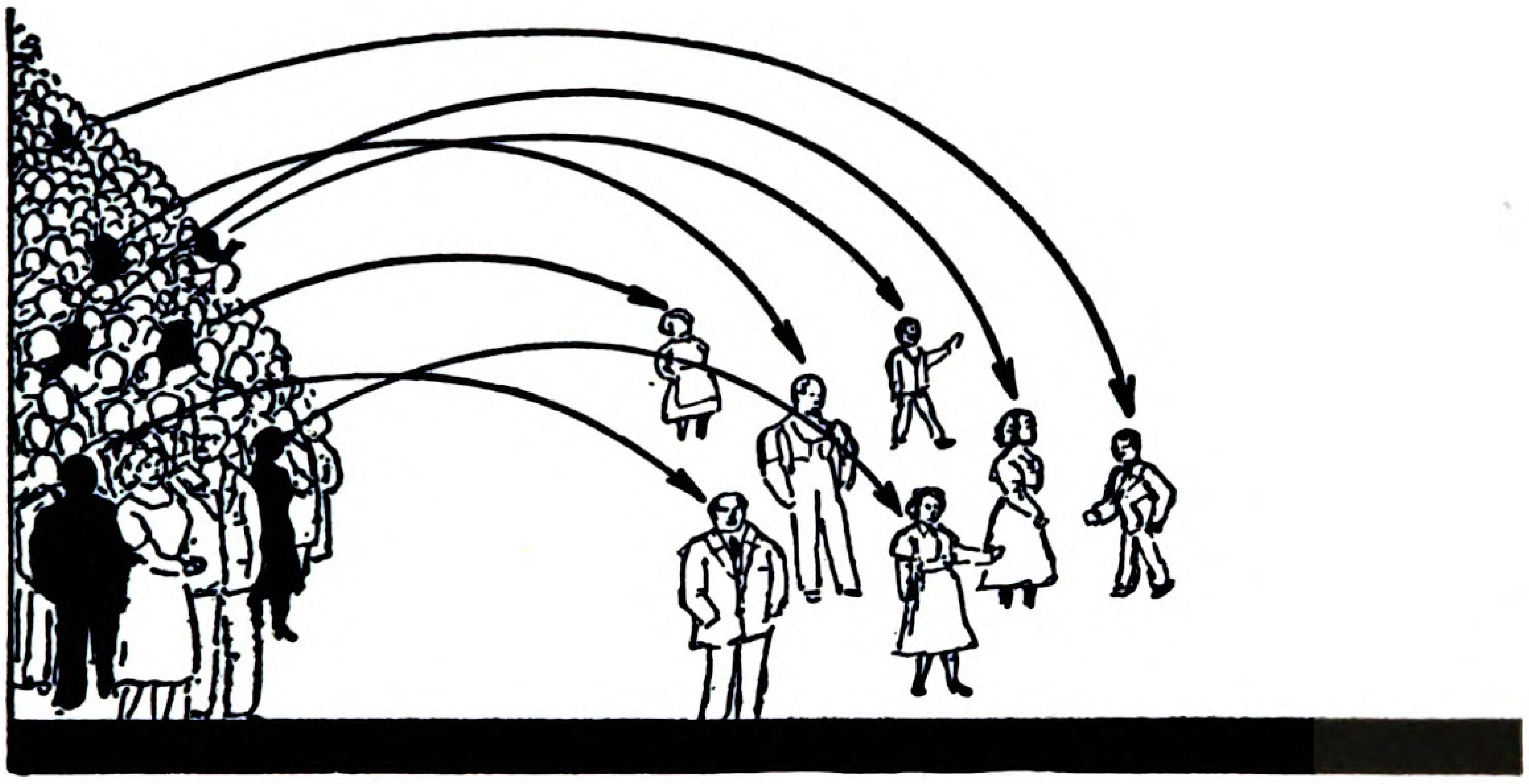
A LARGE SHIPMENT of raw materials arrives at the Alpha Corporation. The receiving clerk opens a few bags and checks part of the contents. He samples the shipment. Since the quality of the sample is satisfactory, he concludes that the entire shipment is satisfactory and may be accepted.

Or: The Alpha Corporation wishes to study what the consumers of their products think about certain features of these products.

It is practically impossible to ask the thousands of people all over the country who have been using Alpha products in the past. However, instead of attempting to ask all the consumers, the Alpha Corporation selects a comparatively small number of them and proceeds to inquire concerning everything the corporation wants to ascertain. This group upon which the investigation is conducted is a sample taken from the total number of Alpha con-

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

sumers, which we call a universe or population. The latter expression is used regardless of whether we deal with persons or objects. Any finite or infinite mass from which we draw a sample, is called a universe, or population.



Our sample has been drawn for the purpose of investigation and we may expect, under certain circumstances, to obtain approximately the same answers as if we had examined all the Alpha consumers.

In our case it is much simpler and considerably less expensive to examine a sample. In many other situations, it may be absolutely impossible to analyze the whole universe and a sample must suffice.

What we are doing is to draw conclusions from a small group and apply them to a large or even

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

infinite group, whose behavior, thinking, etc., are unknown to us. Certainly, risks are involved in making statements about unknown items.

The Alpha Corporation may use different methods of obtaining a satisfactory sample.

First, there is the random sample. It is done in such a way that every consumer of Alpha products has an equal chance of being selected. This requirement sounds simple but it demands much careful planning to guarantee this equal right of every Alpha consumer. No selective factor, no bias must creep in. It would certainly not do to pick all consumers living in nearby areas, or all who have a telephone. These factors would limit the chance of many of the consumers to be included in the sample.

Another method which is sometimes used to select a sample is purposive sampling. Here we select consumers because they are known to have one or more characteristics of the whole group of consumers. The Alpha Corporation may, for instance, have established that the consumers living in the town N are typical for all Alpha consumers. From the standpoint of the Alpha Corporation the consumers of town N can be called "Mr. and Mrs. Consumer." We therefore examine all the consumers of town N as a purposive sample.

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

The products manufactured by the Alpha Corporation are such that they are equally useful to any person in the country. Sex, age, wealth, education, etc., make no difference with regard to Alpha products. The consumers of Alpha products are, therefore, in respect to the use of these products, of the same kind, or homogeneous. We have in that case a homogeneous universe, from which we draw samples.

The Beta Corporation shows a different situation. Beta products differ in importance to the farmer and the city dweller. We shall therefore group the Beta consumers in, say, three groups: people living in rural districts, in cities except metropolitan areas, and people living in metropolitan areas.

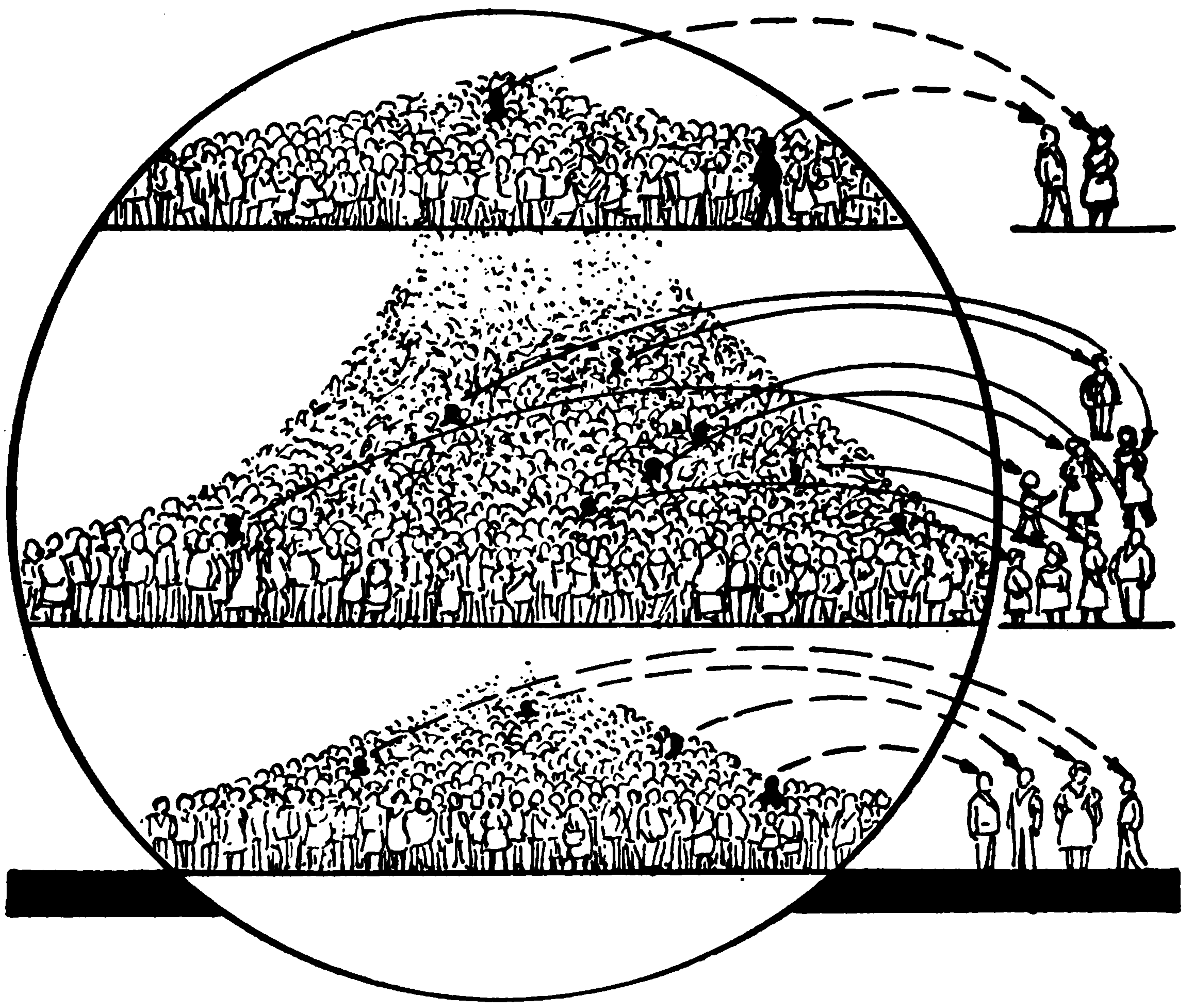
The consumers of Beta products are not of the same kind with regard to Beta products. We were easily able to determine a basis for grouping. The Beta consumers, therefore, are not a homogeneous universe but a heterogeneous one. However, within each of the three groups there is homogeneity with regard to the purpose of our investigation. All the consumers living in rural areas, for instance, are of the same kind. These three groups are called strata.

We may now draw a random sample from each stratum. We shall take into account the size of the

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

stratum and shall draw our sample at random in proportion to the size of each stratum. This type of sample, which is frequently used in market research and opinion surveys, is called a stratified random sample.



We may prefer to select a purposive sample out of the stratified universe. In this case we talk of a stratified purposive sample.

A good sample has to be adequate and reliable.

Digitized by Google
Original from UNIVERSITY OF MICHIGAN

A sample is adequate if a sufficient number of items has been included in the sample. No general rule about the adequate size of the sample can be given since it depends upon several factors, among them, in our case, the variation in the characteristics of the products and the consumers.

Sufficiency of a sample is relative to the precision we wish to obtain in sampling. Greater precision usually means larger samples.

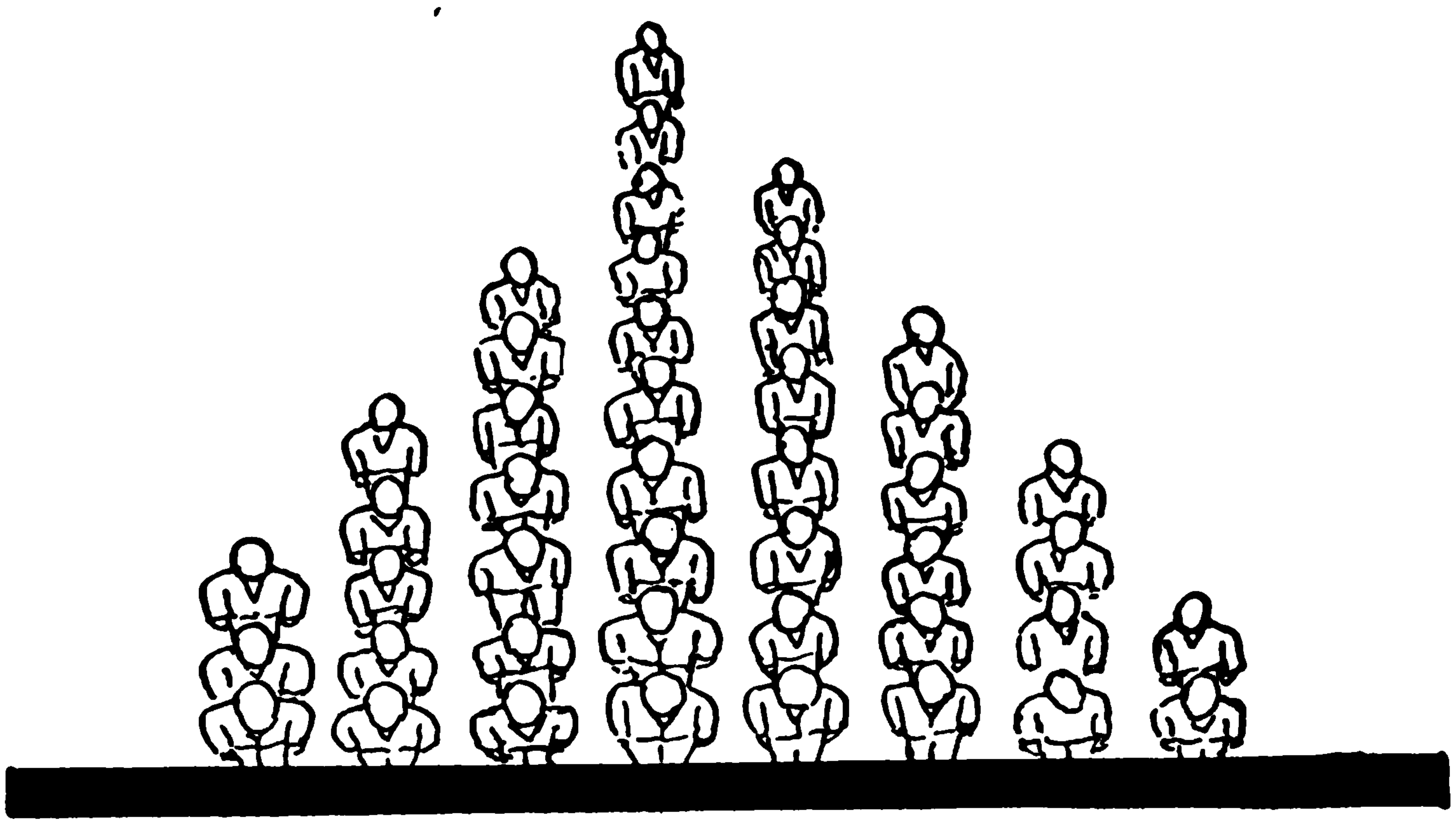
Another question is the reliability of a sample. Would a second and third sample taken from the same universe of consumers show somewhat different results? They probably would. Therefore we must consider the findings based on our sample as just approximations. Statisticians, however, can estimate the limits within which these different samples would vary.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

The Normal Distribution

IN A PREVIOUS INSTANCE, we assembled all of the Alpha workers in order to study their height.

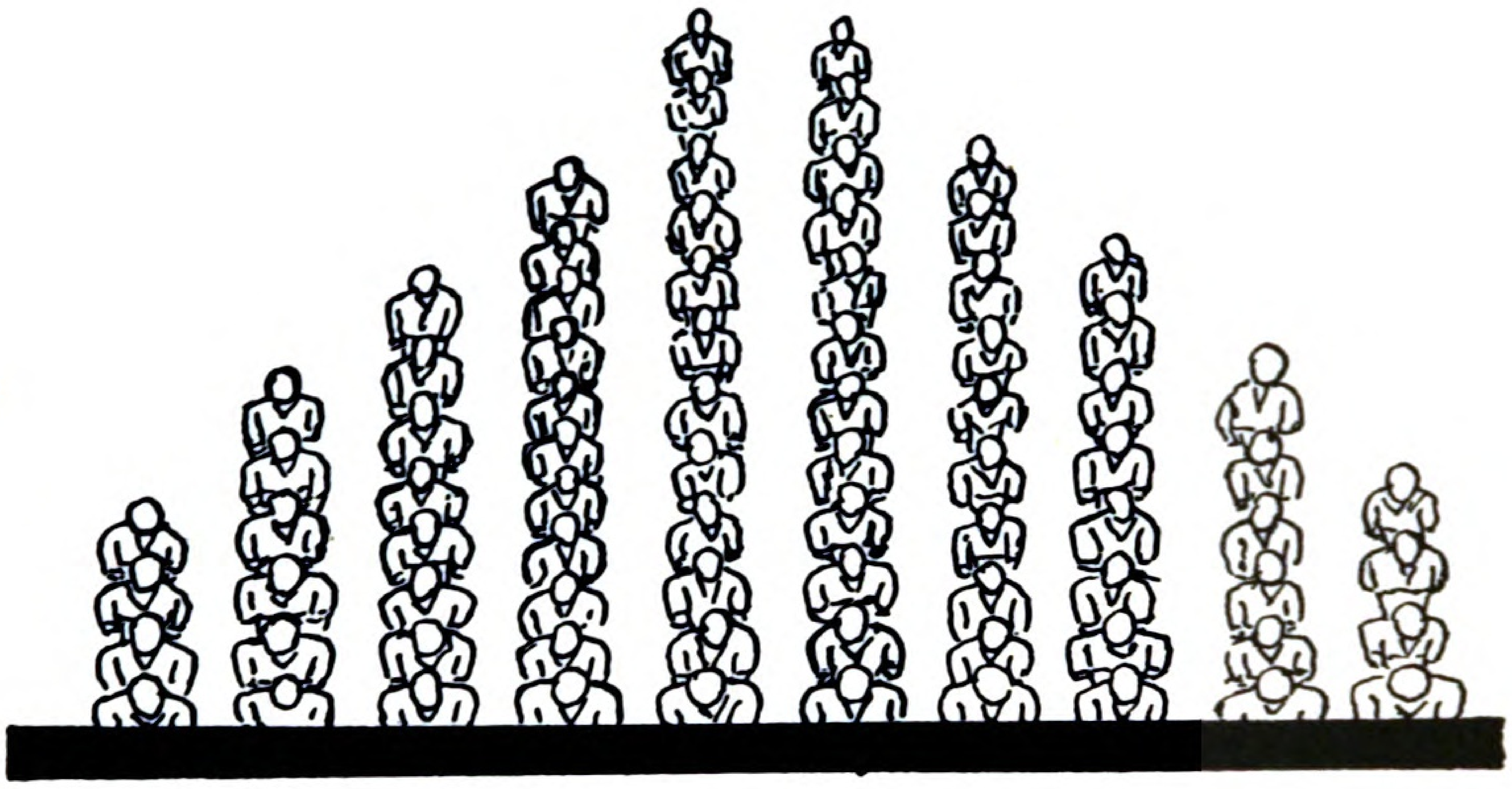


If we could get all male residents of the state together we would see a height distribution that looks more symmetrical.

Finally, the heights of all male inhabitants of the earth may come very close to the ideal symmetrical

Digitized by Google Original from UNIVERSITY OF MICHIGAN

distribution that is referred to as a normal distribution.



This normal curve is a highly important pattern. It is symmetrical in two respects—with regard to skewness and to kurtosis. In fact, the peakedness of any distribution is judged in comparison with the peakedness of the normal distribution.

Normal distribution is the expression of the principle of uniformity in nature which is a correlative of the principle of variation. The height of men varies and still, if we would study the height of all of them, a uniformity would appear. A great number of men have average height; the more a given height departs from the average, the less occurrences of it would be available. We would be

Digitized by Google Original from UNIVERSITY OF MICHIGAN

just as likely to find taller-than-average men as we were to find men who are shorter-than-average. In fact, the deviations in the direction of larger-than-average would be equal in number and magnitude to the deviations in the opposite direction.

The distribution of height data is normal, because the forces that affected the height of every individual have been numerous and of approximately equal importance. That means many forces have caused the height of an individual and none of these forces, like nutrition or climate, had outstanding importance. An effective drug stimulating or retarding growth would have unequal weight as compared with the other forces. Finally, these forces affecting the height of an individual were not connected among themselves, but were independent of one another.

If we increase the number of observed items tremendously, we shall in many cases approach the pattern of a normal distribution, but certainly not in all cases.

Generally, we shall find a normal distribution much more frequently in phenomena of nature than in economic data.

Not only height of human beings but also their weight, foot length, neck size, etc. will be dis-

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

tributed approximately in accordance with the pattern of a normal distribution.

In economic data we shall rarely meet the requirement of approximately equal importance of all causal factors.

Many aspects of the normal distribution have been analyzed thoroughly and the results of these analyses are always applicable to any normal distribution or, since we are thinking in terms of approximations, to any distribution that comes close to this pattern. This applies to the distribution of several economic data.

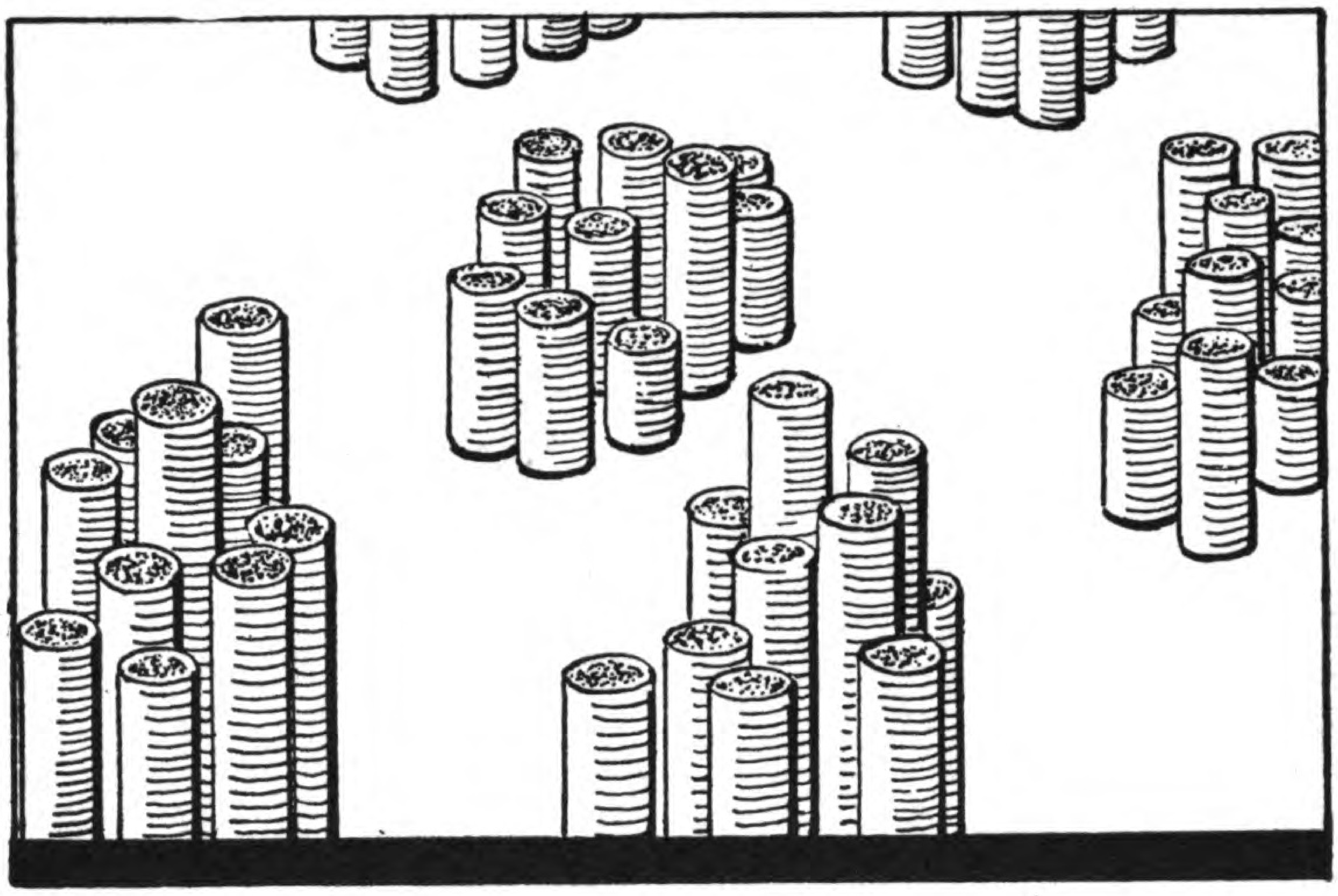
The uses of the concept of a normal distribution are numerous and various. One of the most important uses of this pattern is the following:

The wages of all the workers in the country would not show a normal distribution. Let us see what happens if we draw a great number of random samples from the universe. The universe in this case is wages in the entire country.

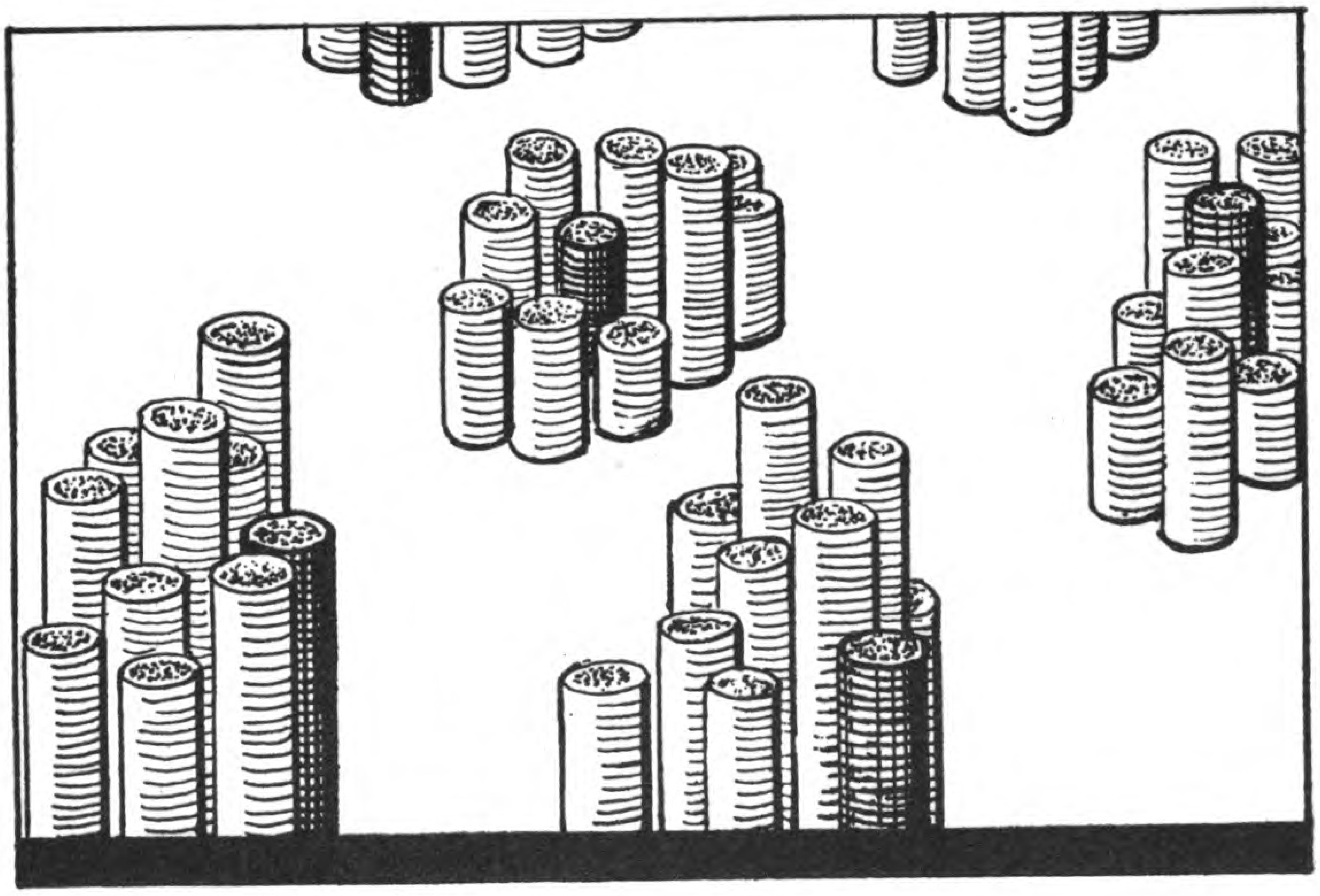
Each sample consists of the wage data of, say, 10 workers.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

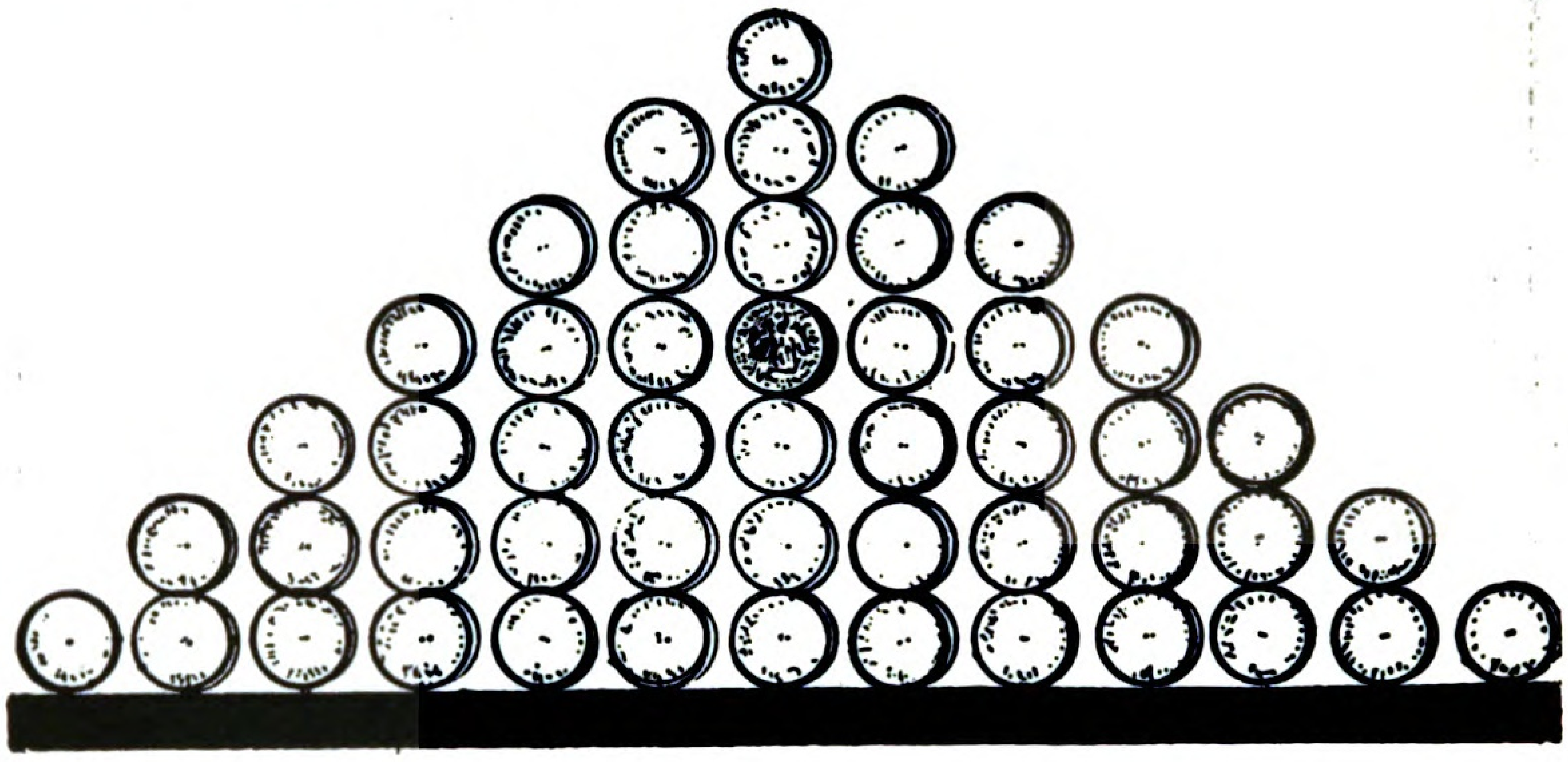


We may find the mean wage of every sample. For simplicity's sake, let us assume that the mean wage really coincides with an existing wage.



Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

If we form a distribution of all these sample means, we shall see that they follow the pattern of a normal distribution.



The mean of these sample means now may be considered the true mean of all wages in the country.

The fact that the normal curve has been analyzed thoroughly has greatly benefited our work with random samples.

Sampling has assumed an increasingly prominent role in statistics and is being further explored and developed. In the future it will certainly be of greatest importance to statistics and its many applications.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

References

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

References

There are only few publications dealing with statistics on a very elementary level. Here are some of them:

ARCO EDITORIAL BOARD. *Statistical Clerk and Junior Statistician.* Arco Publishing Co., New York (Mimeographed).

BARRETT, LAWRENCE A. *Statistics: How to Use and Interpret Them.* Haldeman-Julius Publications, Girard, Kansas.

BUTSCH, R. L. C. *How to Read Statistics.* Bruce Publishing Co., Milwaukee.

CHURCH, DONALD E. *Speed Methods of Statistics.* For Use in Business. Ronald Press Co., New York.

FRANZBLAU, ABRAHAM N. *A Primer of Statistics for Non-Statisticians.* Hebrew Union College, Cincinnati (Mimeographed).

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

GILMAN, STEPHEN. *What the Figures Mean.* Ronald Press Co., New York.

PERGANDE PUBLISHING COMPANY. *Intermediate Statistics.* Milwaukee (Photoprinted).

SCHORLING, RALEIGH, CLARK, JOHN R., AND LANKFORD, FRANCIS G., JR. *Statistics.* World Book Co., Yonkers, N. Y. *Mathematics for the Consumer.* Unit 1: How to Handle Statistics. World Book Co., Yonkers, N. Y.

SUTCLIFFE, WILLIAM G. *Statistics for the Business Man.* Harper, New York.

For further work in statistics college textbooks may be of use. Out of the great number of available textbooks just a few are listed here for purposes of illustration. Many other good textbooks can be found in the bibliographies of the books named below or in the catalogues of libraries.

ARKIN, HERBERT AND COLTON, R. R. *An Outline of Statistical Methods.* Barnes & Noble, New York.

BLAIR, MORRIS M. *Elementary Statistics.* Henry Holt & Co., New York.

BRUMBAUGH, MARTIN A., KELLOGG, LESTER S. AND GRAHAM, IRENE J. *Business Statistics.* R. D. Irwin, Chicago.

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

References

CROXTON, FREDERICK E. AND COWDEN, DUDLEY J. *Practical Business Statistics.* Prentice-Hall, New York.

EASTWOOD, R. PARKER. *Sales Control by Quantitative Methods.* Columbia University Press, New York.

MILLS, FREDERICK C. *Statistical Methods; Applied to Economics and Business.* Henry Holt & Co., New York.

NEISWANGER, WILLIAM A. *Elementary Statistical Methods.* Macmillan Co., New York.

WALKER, HELEN M. *Elementary Statistical Methods.* Henry Holt & Co., New York.

For definitions of statistical terms, the reader of statistical presentations may find the following dictionary of help:

KURTZ, ALBERT K. AND EDGERTON, HAROLD A. *Statistical Dictionary of Terms and Symbols.* John Wiley & Sons, New York.

Digitized by Google | Original from UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Index

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Index

Aggregative Price Index, 66
Arithmetic Mean—*see* "Mean"
Association, 77, 82, 87
Average, 29, 30, 33, 34, 70
Average Deviation, 36, 39, 43—
see also "Mean Deviation"
Average-of-Relatives Index, 68

Bureau of Labor Statistics, 74, 75, 76

Causal Relationship, 80
Coefficient of Correlation, 85
Coefficient of Variation, 41
Consumers' Price Index, 75, 76
Corn-Hog Ratio, 20
Correlation, 77-87
Cost of Living Index, 75, 76
Current Ratio, 19
Curvilinear Correlation, 84
Curvilinear Trend, 52
Cycle, 56, 57, 58, 61

Deflation, 73
Deseasonalized Data, 58, 59
Dispersion, 34, 38

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Federal Reserve Index, 75, 76
Fixed Assets Ratio, 20
Forecasting, 52, 56
Frequency Distribution, 25-47
Functional Relationship, 80

Heterogeneous, 92
Homogeneous, 92

Index Numbers, 63-76
Interquartile Range, 35
Irregular Fluctuation, 58, 61

Kurtosis, 47, 96

Linear Correlation, 83

Mean, 29, 30, 31, 33, 38, 40, 43, 99
Mean Deviation, 36, 37—*see also* "Average Deviation"
Median, 28, 30, 31, 32, 43
Mode, 27, 30, 31, 43
Multiple Correlation, 86

Negative Correlation, 77, 83
Net Worth Ratio, 20
Normal Distribution, 95-100
Normal Ratios, 19

Partial Correlation, 87
Per Capita Ratio, 19
Perfect Correlation, 83
Population, 90
Positive Correlation, 77, 82, 83
Purposive Sample, 91

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Quartile, 32, 35
Quartile Deviation, 35, 43, 44

Random Sample, 91, 98, 100
Range, 34, 35, 43
Ratios, 17-23
Real Sales, 73
Real Wages, 74
Regression Line, 82, 83

Safe Ratios, 19
Sampling, 70, 89-94
Seasonal Pattern, 55
Seasonal Variation, 53, 58, 59, 61, 70, 76
Semi-Interquartile Range, 35
Simple Correlation, 84, 86
Skewness, 46, 96
Standard Deviation, 38, 39, 40, 44
Stratum, 92, 93

Time Series, 49-61
Trend, 51, 52, 53, 58, 59, 60, 61

Universe, 90, 98

Variance, 38
Variation, 13, 34, 94

Weighted Aggregative Index, 68, 74, 75
Weighted Average-of-Relatives Index, 75
Wholesale Price Index, 74, 76

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google

Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN

Digitized by Google
Original from
UNIVERSITY OF MICHIGAN